

データサイエンティスト スキルチェックリスト

「データサイエンティスト スキルチェックリスト」は、データサイエンティストに必要とされるスキルをチェックリスト化したものです。チェックリスト内の各項目のスキルレベルは、★の数によって示し、見習いレベル（Assistant Data Scientist）～ 標準レベル（Full Data Scientist）までを判定します。

スキルレベル	判定基準
① Senior Data Scientist（業界を代表するレベル） ★★★★★	—
② Full Data Scientist（標準レベル） ★★★	★★★の全項目のうち、50%を満たしている。
③ Associate Data Scientist（独り立ちレベル）★★	★★の全項目のうち、60%を満たしている。
④ Assistant Data Scientist（見習いレベル）★	★の全項目のうち、70%を満たしている。

※「必須スキル」に○がついている項目は、判定基準を満たしていても、この項目が達成されていないとそのレベルとは認められない項目として設定しています。

※ 独り立ちレベル以上のレベルは、下位のレベルを満たしていることが前提となります。

*引用・改変時の注意事項

- ・チェックリストを引用される場合「一般社団法人データサイエンティスト協会 スキルチェックリストより引用」と引用元を明示願います。その他については著作権法に従っての引用を願います。
- ・チェックリストを改変された場合「一般社団法人データサイエンティスト協会 スキルチェックリストを改変」と改変した旨、明示願います。

スキルカテゴリー一覧							
サブカテゴリ		項目数	サブカテゴリ		項目数		
データサイエンスカ	1	統計数理基礎	16	データエンジニアリングカ	1	環境構築	21
	2	予測	17		2	データ収集	16
	3	検定/判断	11		3	データ構造	11
	4	グルーピング	14		4	データ蓄積	17
	5	性質・関係性の把握	14		5	データ加工	13
	6	サンプリング	5		6	データ共有	14
	7	データ加工	8		7	プログラミング	22
	8	データ可視化	37		8	ITセキュリティ	15
	9	分析プロセス	5	項目数		129	
	10	データの理解・検証	23	ビジネスカ	1	行動規範	12
	11	意味合いの抽出、洞察	4		2	論理的思考	18
	12	機械学習	20		3	プロジェクトプロセス	20
	13	時系列分析	7		4	データ入手	4
	14	言語処理	13		5	データの理解・検証	3
	15	画像・動画処理	8		6	意味合いの抽出、洞察	5
	16	音声/音楽処理	5		7	解決	4
	17	パターン発見	3		8	事業に実装する	8
	18	グラフィカルモデル	3		9	活動マネジメント	20
	19	シミュレーション/データ同化	5		10	知財	6
	20	最適化	10	項目数		100	
項目数		228	項目数合計			457	

ver 2.00 (as of 10/23/2017)

一般社団法人データサイエンティスト協会 Copyright © 2017 The Japan DataScientist Society. All Rights Reserved.

スキルチェックリスト 2017年 改訂版 <ビジネスカ>

▼色の付帯のスキル

NO	Sub No	スキルカテゴリ	スキルレベル	サブカテゴリ	チェック項目	DE	DS	必須スキル
1	1	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つき行動できる			○
2	2	行動規範	★★	ビジネスマインド	ビジネスではスピード感がより重要であることを認識し、時間と情報が限られた状況下でも、言わば「ザックリ感」を持って素早く意思決定を行うことができる			○
3	3	行動規範	★★	ビジネスマインド	作業ありきではなく、本質的な問題（イシュー）ありきで行動できる			○
4	4	行動規範	★★	ビジネスマインド	分析で価値ある結果を出すためにはしばしば仮説検証の繰り返しが必要であることを理解し、粘り強くタスクを完遂できる			○
5	5	行動規範	★★★	ビジネスマインド	プロフェッショナルとして、作業量ではなく、生み出す価値視点で常に判断、行動でき、依頼元にとって真に価値あるアウトプットを生み出すことをコミットできる			○
6	6	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付けている（データのねつ造、改ざん、濫用を行わないなど）			○
7	7	行動規範	★★	データ倫理	チーム全員がデータを取り扱う人間として相応しい倫理を持てるよう、適切にチームを管理できる			
8	8	行動規範	★★★	データ倫理	データの取り扱いに関する、会社や組織全体の倫理を維持、向上させるために、必要な制度や仕組みを策定し、その運営を主導することができる			
9	9	行動規範	★	法令	個人情報に関する法令の概要を理解し、守るべきポイントを説明できる			○
10	10	行動規範	★★	法令	担当するビジネスや業界に関係する法令を理解しており、データの保持期間や運用ルールに活かすことができる	*		○
11	11	行動規範	★★	法令	個人情報の扱いに関する法令、その他のプライバシーの問題、依頼元との契約約款に基づき、匿名化すべきデータを選別できる（名寄せにより個人を特定できるもの、依頼元がデータ処理の結果をどのように保持し利用するのかなども考慮して）			
12	12	行動規範	★★	法令	匿名加工情報について理解しており、適切な方法で匿名加工情報を扱うことができる	*	*	
13	1	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			○
14	2	論理的思考	★★	MECE	初見の領域に対して、抜け漏れや重複をなくすることができる			
15	3	論理的思考	★★★	MECE	未知の領域であっても、類似する事象の推測などを活用し、抜け漏れや重複をなくすることができる			
16	4	論理的思考	★★	データをもとに論理的に捉える	様々なデータや事象を、階層やグルーピングによって、構造化できる（ピラミッド構造）			○
17	5	論理的思考	★★	データをもとに論理的に捉える	仮説思考を用いて、論点毎に分析すべき点を識別できる			
18	6	論理的思考	★★★	データをもとに論理的に捉える	論理的な整理にとらわれず、批判的・複合的な視点で課題を識別できる			
19	7	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			○
20	8	論理的思考	★★	言語化能力	通常見受けられない現象の場合においても、分析結果の意味合いを既知の表現を組み合わせ、言語化できる			
21	9	論理的思考	★★★	言語化能力	データ表現に適した言葉がない場合でも、共通認識が形成できるような言葉を新たに作り出すことができる			
22	10	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）			
23	11	論理的思考	★★	ストーリーライン	因果関係に基づいて、ストーリーラインを作る（脱線→気づき→打ち手、So What?, Why So?など）			
24	12	論理的思考	★★★	ストーリーライン	相手や内容に応じて、自在にストーリーラインを組み上げることができる			

NO	Sub No	スキルカテゴリ	スキルレベル	サブカテゴリ	フェック項目	DE	DS	必須スキル
25	13	論理的思考	★	ドキュメンテーション	1つの図表～数枚程度のドキュメントを論理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、興味合い、ネクストステップ）			
26	14	論理的思考	★★	ドキュメンテーション	10～20枚程度のミニパッケージ（テキスト&図表）、もしくは5ページ程度の図表込みのビジネスレポートを論理立てて作成できる			
27	15	論理的思考	★★★	ドキュメンテーション	30～50枚程度のフルパッケージ（テキスト&図表）、もしくは10ページ以上のビジネスレポートを論理立てて作成できる			
28	16	論理的思考	★	説明能力	報告に対する論拠不足や論理破綻を指摘された際に、相手の主張をすみやかに理解できる			○
29	17	論理的思考	★★	説明能力	論理的なプレゼンテーションができる			○
30	18	論理的思考	★★★	説明能力	プレゼンテーションの相手からの質問や反論に対して、説得力のある形で回答できる			○
31	1	プロジェクトプロセス	★	基礎知識	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			○
32	2	プロジェクトプロセス	★	基礎知識	課題や仮説を言語化することの重要性を理解している			○
33	3	プロジェクトプロセス	★	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			○
34	4	プロジェクトプロセス	★★	KPI	自らが関連する事業領域であれば、複数の課題レイヤーにまたがっていても、KPIを整理・構造化できる			
35	5	プロジェクトプロセス	★★★	KPI	初見の事業領域であっても、KPIを構造化し、重要なKPIを見極められる			
36	6	プロジェクトプロセス	★	課題設定の前提理解	自分の担当する業界について、市場規模、主要なプレーヤー、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			
37	7	プロジェクトプロセス	★★	課題設定の前提理解	事業モデルやバリューチェーンなどの特徴や事業の主たる課題を自力で構造的に理解でき、問題の大本を整理できる			
38	8	プロジェクトプロセス	★★★	課題設定の前提理解	領域の主要課題を他領域の課題との関連も含めて構造的に理解でき、問題の大本を定義できる			
39	9	プロジェクトプロセス	★	課題の定義	担当する業務領域であれば、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）			○
40	10	プロジェクトプロセス	★★	課題の定義	取り扱う課題領域に対して、新規課題の切り分けや枠組み化ができる			
41	11	プロジェクトプロセス	★★★	課題の定義	仮説や可視化された問題がなくとも、解くべき課題を構造的に整理でき、見極めるべき論点を特定できる			
42	12	プロジェクトプロセス	★	プロジェクト推進	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	*		
43	13	プロジェクトプロセス	★★	プロジェクト推進	財務会計と管理会計の大まかな枠組みを理解し、必要に応じて分析プロジェクト設計ができる			
44	14	プロジェクトプロセス	★★	プロジェクト推進	アジャイル開発体制のポイントを理解した上で、アジャイルな開発チームを立ち上げ、推進することができる	*		
45	15	プロジェクトプロセス	★★	プロジェクト推進	類似事例の実績やPoC（Proof of Concept）を適宜利用して、プロジェクト計画に関わるステークホルダー間の合意を形成できる			
46	16	プロジェクトプロセス	★★★	プロジェクト推進	期待される成果が達成できないケースを早期に見極め、プロジェクトの終了条件をステークホルダーと整理・合意できる			
47	17	プロジェクトプロセス	★★	アプローチ設計	最終的な結論に関わる部分や、ストーリーラインの骨格に大きな影響を持つ部分から着手するなど、取り進むべき分析上のタスクの優先度を判断できる	*		○
48	18	プロジェクトプロセス	★★★	アプローチ設計	他社による模倣を防ぐなど、競争力を保つ観点でアプローチの設計ができる（機械化や標準化の範囲を絞るなど）			
49	19	プロジェクトプロセス	★★★	分析価値の判断	プロジェクトの開始時点で、入手可能なデータ、分析手法、インフラ、ツールの生み出すビジネス価値の見積りをまとめることができる	*	*	
50	20	プロジェクトプロセス	★★★	プロジェクトの評価	分析結果が当初の目的を満たしていない場合に、必要に応じてプロジェクト全体を再設計できる	*	*	○

NO	Sub No	スキルカテゴリ	スキルレベル	サブカテゴリ	チェック項目	DE	DS	必須スキル
51	1	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、入手できる	*	*	○
52	2	データ入手	★	データ入手	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	*		
53	3	データ入手	★★	データ入手	自身が担当するプロジェクトやサービスを越えて、必要なデータのあたりをつけることができる	*	*	
54	4	データ入手	★★★	データ入手	組織全体及び関連する社外のデータを見直し、必要なデータのあたりをつけることができる	*	*	
55	1	データの理解・検証	★	ビジネス視点	ビジネス視点で仮説を持ってデータをみることの重要性と、仮説と異なる結果となったときにそれが重大な知見である可能性を理解している			○
56	2	データの理解・検証	★★	ビジネス視点	統計手法による間値に対してビジネス視点で納得感のある調査ができる（年齢の刻み、商品単価、購入周期を考慮した単変数のカテゴリーなど）		*	
57	3	データの理解・検証	★★★	ビジネス視点	分析プロセス全体を通して常時、ビジネス視点での妥当性をチェックし、データから得られた示唆が価値ある知見であるか判断できる			
58	1	意味合いの抽出・洞察	★★★	開示	担当および関連する分析プロジェクトのデータ、分析結果を顧客、外部に開示すべきか判断できる			
59	2	意味合いの抽出・洞察	★★	洞察	ビジネス上の意味を捉えるために、特異点、相違性、傾向性、関連性を見出した上で、ドメイン知識を持つ人に適切な質問を投げかけられる		*	
60	3	意味合いの抽出・洞察	★★	洞察	データの特徴を見て意味合いの明確化に向けた分析の深掘り、データ見直しの方向性を設計できる	*		
61	4	意味合いの抽出・洞察	★★	洞察	分析結果を眺め、起きている事象の背景や意味合い（真実）を見極めることができる	*		
62	5	意味合いの抽出・洞察	★★	評価	担当する分析プロジェクトの分析結果を見て検討目的と合っているか再評価できる		*	○
63	1	解決	★★	総合的評価	適用される業務内容に応じて、モデルの総合的な評価ができる		*	
64	2	解決	★★	提案	分析的検討に基づき、担当業務に対する必要なアクション、改革案を整理して結論を導くことができる			
65	3	解決	★★★	提案	分析的検討に基づき、経営レベルで必要なアクション、改革案を整理して結論を導くことができる			○
66	4	解決	★★★	提案	データを利活用した持続性のある事業モデルを設計できる			
67	1	事業に実装する	★★	実装	現場に実装する際、現場での実行可能性を考慮し適切に対応できる	*		
68	2	事業に実装する	★★	実装	担当する案件が予算内で解決するように取り組みをデザインし、現場に実装できる	*		
69	3	事業に実装する	★★★	実装	費用対効果、実行可能性、業務負荷を考慮し事業に実装ができる	*		
70	4	事業に実装する	★★	異分野とのコミュニケーション	異なるスキル分野の専門家や事業者と適切なコミュニケーションをとりながら事業・現場への実装を進めることができる			○
71	5	事業に実装する	★	モニタリング	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			○
72	6	事業に実装する	★★	モニタリング	事業・現場へ実装するにあたりモニタリングの仕組みを適切に組み込むことができる	*	*	
73	7	事業に実装する	★★★	モニタリング	既存のPDCAに新たな仕組みを行い、次の改善的な取り組みにつなげることができる			
74	8	事業に実装する	★★★	横展開	特定のビジネス課題に向けた新しいソリューションを個別の現場の特性を考慮し横展開できる	*	*	
75	1	活動マネジメント	★★★	データ倫理	マルウェア、DDoS攻撃などの深刻なセキュリティ攻撃を受けた場合に対応する最新の技術を把握し、対応する専門組織（CSIRT）の構成を責任者にすみやかに提案できる	*		
76	2	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了できる			○

NO	Sub No	スキルカテゴリ	スキルレベル	サブカテゴリ	チェック項目	DE	DS	必須スキル
77	3	活動マネジメント	★★	リソースマネジメント	自身とチームメンバーのスキルを把握し、適切なプロジェクト管理ができる			○
78	4	活動マネジメント	★★	リソースマネジメント	担当するプロジェクトで、設定された予備ツール、システム環境を適切に活用し、プロジェクト進行できる	*	*	
79	5	活動マネジメント	★★	リソースマネジメント	5名前後のチームを設計し、スケジュール通りに進行させ、ステークホルダーに対して期待値に見合うアウトプットを安定的に生み出せる			
80	6	活動マネジメント	★★★	リソースマネジメント	プロジェクトに求められるスキル要件と各メンバーのスキル・成長目標・性格をふまえ、現実的にトレードオフ解消とシナジーを狙ったリソースマネジメントができる	*	*	
81	7	活動マネジメント	★★★	リソースマネジメント	プロジェクトメンバーの技量を把握した上で、プロジェクト完遂に必要なツール選定、予算策定、スコープ設定、またはアウトソーシング体制を検討・構築できる	*	*	
82	8	活動マネジメント	★★★	リソースマネジメント	複数のチームを設計し（総合的なプロジェクトマネジメント）、スケジュール通りに進行させ、複合的なステークホルダーに対し、期待値を超えたアウトプットを安定的に生み出せる	*	*	
83	9	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			○
84	10	活動マネジメント	★★	リスクマネジメント	担当するプロジェクトでの遅延や障害などの発生を検知し、リカバリーするための提案・設計ができる			
85	11	活動マネジメント	★★★	リスクマネジメント	プロジェクトに何らかの遅延・障害などが発生した場合、適切なリカバリー手順の判断、リカバリー体制構築、プロジェクトオーナーに対する迅速な対応ができる			○
86	12	活動マネジメント	★★	育成/ナレッジ共有	自身とチームメンバーのスキルを大まかに把握し、担当するプロジェクトを通してチームメンバーへのスキルアドバイスやスキル成長のための目標管理ができる	*	*	
87	13	活動マネジメント	★★	育成/ナレッジ共有	チームメンバーのスキルに応じ、研修参加や情報収集への適切なアドバイスやチーム内でのナレッジ共有を推進できる	*	*	
88	14	活動マネジメント	★★★	育成/ナレッジ共有	チームの各メンバーに対し、データサイエンティストとしてのスキル目標の設定、到達させるためのプロジェクトを通じた適切なアドバイスができる	*	*	
89	15	活動マネジメント	★★★	育成/ナレッジ共有	チーム育成の上で、データサイエンティストに求められるスキルについて、研修制度の設計やナレッジ共有の仕組み構築と運営ができる	*	*	
90	16	活動マネジメント	★★★	育成/ナレッジ共有	チームに必要な情報やデータサイエンスの新しい技術・手法に関する情報収集戦略やラーニング方法を主導し、自ら情報を取捨選択し、チームにフィードバックできる	*	*	
91	17	活動マネジメント	★★	営業マネジメント	ビジネス要件を整理し、分析・データ活用プロジェクトを企画・提案することができる	*	*	○
92	18	活動マネジメント	★★★	営業マネジメント	依頼元やステークホルダーのビジネスをデータ面から理解し、分析・データ活用プロジェクトを立ち上げ、プロジェクトにかかるコストと依頼元の利益を説明できる	*	*	
93	19	活動マネジメント	★★	組織マネジメント	データサイエンスチームの役割を認識し、担当するプロジェクトにおいて、組織内や他部門・他社間でのタスク設定や調整ができる			
94	20	活動マネジメント	★★★	組織マネジメント	データサイエンスチームを自社・他社の様々な組織と関連付け、対象組織内での役割の規定、目標設定を行うことができる			○
95	1	知財	★★	AI・モデル開発	既存ライブラリなどを利用し、解析または開発を推進する際に、知財リスクの確認など、適切な対応をとることができる			○
96	2	知財	★★★	AI・モデル開発	独自に開発する手法・アルゴリズム構築の際に、適切な専門家の助力を得て知財リスクの管理を行うことができる			
97	3	知財	★★★	AI・モデル開発	独自に開発した手法・アルゴリズムに対する他者からの権利侵害に備え、特許出願を含む適切な対応を行うことができる			
98	4	知財	★★★	AI・モデル開発	AI・モデル開発において、事業・現場に即したガイドラインを定義できる	*	*	
99	5	知財	★★★	AI・モデル開発	AI・モデルの活用・責任範囲に関し事業・現場に即したガイドラインを定義できる	*	*	

NO	Sub No	スキルカテゴリ	スキルレベル	サブカテゴリ	チェック項目	DE	DS	必須スキル
100	5	知財	★★★	AI・モデル開発	学習済みモデルや生成されたデータの価値を理解した上で権利関係を明確にし、利用許諾など契約の枠組みを定義できる	*	*	

スキルチェックリスト 2017年 改訂版 <データサイエンスカ>

▼ 統計分野のスキル

NO	Sub No	スキルカテゴリ	スキルレベル	サブカテゴリ	チェック項目	BZ	DE	必須スキル
1	1	統計数理基礎	★	統計数理基礎	1+4+9+16+25+36を2を用いて表せる			○
2	2	統計数理基礎	★	統計数理基礎	$y=\log_2(x)$ の逆関数を説明できる			○
3	3	統計数理基礎	★	統計数理基礎	階列や組合せを式 nPr , nCr を用いて計算できる			○
4	4	統計数理基礎	★	統計数理基礎	条件付き確率の意味を説明できる			○
5	5	統計数理基礎	★	統計数理基礎	平均(相加重平均)、中央値、最頻値の算出方法の違いを説明できる			○
6	6	統計数理基礎	★	統計数理基礎	母集団データ(3,4,5,5,7,8)の分散と標準偏差を電卓を用いて計算できる			○
7	7	統計数理基礎	★	統計数理基礎	母(集団)平均が標本平均とは異なることを説明できる			○
8	8	統計数理基礎	★	統計数理基礎	標準正規分布の分散と平均がいくつかわかる			○
9	9	統計数理基礎	★	統計数理基礎	相関関係と因果関係の違いを説明できる			○
10	10	統計数理基礎	★	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			○
11	11	統計数理基礎	★	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			○
12	12	統計数理基礎	★	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			○
13	13	統計数理基礎	★	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中心極限定理により正規分布に近似されることを知っている			○
14	14	統計数理基礎	★	統計数理基礎	変数が量的、質的どちらの場合の関係の強さも算出できる			○
15	15	統計数理基礎	★	統計数理基礎	ベイズの定理を説明できる			○
16	16	統計数理基礎	★★	統計数理基礎	ベイズ統計と頻度論による従来の統計との違いを、尤度、事前確率、事後確率などの用語を用いて説明できる			○
17	1	予測	★	予測	単回帰分析について最小二乗法、回帰係数、標準誤差の説明ができる			○
18	2	予測	★★	予測	重回帰分析において偏回帰係数と標準偏回帰係数、重相関係数について説明できる			○
19	3	予測	★★	予測	重回帰や判別を実行する際に変数選択手法の特徴を理解し、適用できる			○
20	4	予測	★★	予測	ニューラルネットワークの基本的な考え方を理解し、出力される「ダイアグラム」の入力層、隠れ層、出力層の概要を説明できる			○
21	5	予測	★★	予測	重回帰分析において多重共線性の対応ができ、適切に変数を評価・除去して予測モデルが構築できる			○
22	6	予測	★★	予測	決定木分析においてCHAID、C5.0などのデータ分割のアルゴリズムの特徴を理解し、適切な方式を選定できる			○
23	7	予測	★★	予測	線形回帰分析が量的な変数を予測するのに対して、ロジスティック回帰分析は何を予測する手法か(発生確率予測など)を説明でき、実際に使用できる			○
24	8	予測	★★★	予測	予測アルゴリズムに応じ、目的変数と説明変数に対する必要な変数加工処理を設計、実施できる			○
25	9	予測	★★★	予測	予測モデル構築において頑健性を維持するための具体的な方法を設計、実施できる			○
26	10	予測	★★★	予測	尤度と最尤推定についての説明ができる(尤度関数、ネイマンの分解定理、順序統計量)			○

NO	Sub No	スキルカテゴリ	スキルレベル	サブカテゴリ	チェック項目	BZ	DE	必須スキル
27	11	予測	★★★	予測	予測対象データの分布をみて、分布形状に適合した計算法の非線形回帰モデルを構築できる			
28	12	予測	★★★	予測	ロジスティック回帰分析において回帰パラメータとオッズ比の関係について説明できる			
29	13	予測	★★★	予測	ロジスティック回帰分析を行う際に、最小2乗法ではなく最尤法を使う際の利点（回帰誤差が近似的に正規分布しなくても適用できるなど）を説明し、適用することができる			
30	14	予測	★★★	予測	目的（予測・真のモデル推定など）に応じて、適切な損失関数とモデル選択基準（AIC（赤池情報量規準）、BIC（ベイズ情報量規準）、MDL（最小記述長）など）を選択し、モデル評価ができる			
31	15	予測	★★★	予測	データと分析要件から、モデル精度のモニタリング設計・実施と劣化が見込まれるモデルに対するリモデルの設計ができる			
32	16	予測	★★	予測	分析要件に応じ、量的予測のためのモデリング手法（重回帰、決定木、ニューラルネットワークなど）の選択とパラメータ設定、結果の評価、チューニングを適切に設計・実施・指示できる			○
33	17	予測	★★★	予測	当該分野に則したベイズ統計に基づくアルゴリズムを理解し、モデルを構築できる			
34	1	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			
35	2	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			○
36	3	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			
37	4	検定/判断	★	検定/判断	片側検定と両側検定の違いを説明できる			
38	5	検定/判断	★	検定/判断	対応のあるデータと対応のないデータの違いを説明できる			
39	6	検定/判断	★	検定/判断	2群の平均値の差の検定手法を知っている			
40	7	検定/判断	★★	検定/判断	t検定を理解して、パラメトリックな2群の検定を活用することができる			○
41	8	検定/判断	★★	検定/判断	様々な分散分析の考え方（一元配置、多重比較、二元配置）を理解して、パラメトリックな多群の検定を活用することができる			
42	9	検定/判断	★★	検定/判断	ワイルコクソンU検定（マンホイットニーのU検定）を理解して、ノンパラメトリックな2群の検定を活用することができる			
43	10	検定/判断	★★	検定/判断	クラスカル・ウォリス検定を理解して、ノンパラメトリックな多群の検定を活用することができる			
44	11	検定/判断	★★	検定/判断	カイ二乗検定、フィッシャーの直接確率検定を理解して、分割表における群間の関連性の検定を活用することができる			
45	1	グルーピング	★	グルーピング	教師あり学習の分類モデルと教師なし学習のグループ化の違いを説明できる			○
46	2	グルーピング	★	グルーピング	判別分析とクラスター分析の概要や使い方を説明できる			
47	3	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			
48	4	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がわかり、適切に解釈できる			
49	5	グルーピング	★★	グルーピング	非階層クラスター分析において、分析目的に合致したクラスター数を決定することができる			
50	6	グルーピング	★★	グルーピング	階層クラスター分析における代表的なクラスター間距離（群平均法、Ward法、最長一致法など）の概念を理解し、目的に合致した最適な手法で分析できる			○
51	7	グルーピング	★★	グルーピング	各種距離関数（ユークリッド距離、マンハッタン距離、cos類似度など）を理解し、目的に合致した最適な手法で分析できる			

NO	Sub No	スキルカテゴリ	スキルレベル	サブカテゴリ	チェック項目	BC	DE	必須スキル
52	8	グルーピング	★★	グルーピング	分析対象に応じて、線形の判別分析モデルと非線形の判別分析モデルを適切に使い分けることができる			
53	9	グルーピング	★★	グルーピング	k-means法は局所最適解であるため初期値問題があることを理解し、適切な初期値を設定して分析を行える			○
54	10	グルーピング	★★★	グルーピング	距離の公理を知っており、距離の公理を満たさない場合（[1-cos類似度]など）のクラスター分析を適切に行える			
55	11	グルーピング	★★★	グルーピング	k-meansの派生手法（x-means、k-means++、ファジィk-meansなど）を理解し、目的に合致した最適な手法で分析できる			
56	12	グルーピング	★★★	グルーピング	k-meansとカーネルk-means（非線形クラスタリング）、スペクトラルクラスタリングの違いを理解し、試行の中で最適な手法を選択・実行できる			
57	13	グルーピング	★★★	グルーピング	自己組織化マップ（SOM）、Affinity Propagation、混合分布モデル、ディリクレ過程混合モデルを理解し、試行の中で最適な手法を選択・実行できる			
58	14	グルーピング	★★★	グルーピング	反復子ニューラルネットワーク（オートエンコーダ）、One-class SVM（Support Vector Machine）、マハラノビス距離を用いた異常検知の手法を理解し、試行の中で最適な手法を選択・実行できる			
59	1	性質・関係性の把握	★	性質・関係性の把握	適切なデータ区間設定でヒストグラムを作成し、データのバラつき方を把握できる			○
60	2	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			○
61	3	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することができる			○
62	4	性質・関係性の把握	★★	性質・関係性の把握	多重（質問間）クロス表などを駆使して、データから適切なインサイトを得ることができる	*		○
63	5	性質・関係性の把握	★★	性質・関係性の把握	主成分分析と因子分析の違いや使い分けを説明できる			○
64	6	性質・関係性の把握	★★	性質・関係性の把握	正準相関分析を説明、活用できる			
65	7	性質・関係性の把握	★★	性質・関係性の把握	分析の目的と用いるデータの種類から、正規分布を前提とした多変量解析の適切な手法を選択できる			
66	8	性質・関係性の把握	★★	性質・関係性の把握	因子分析における、因子負荷量や因子軸の回転について説明できる			
67	9	性質・関係性の把握	★★	性質・関係性の把握	コレスポンデンス（対応）分析と数量化3類の類似点と違いを説明出来る			
68	10	性質・関係性の把握	★★	性質・関係性の把握	コンジョイント分析を用いて効用値と寄与率のグラフを描くことができる			
69	11	性質・関係性の把握	★★	性質・関係性の把握	適切な類似度を設定した上で、多次元尺度構成法を用いてポジショニングマップを描くことができる			
70	12	性質・関係性の把握	★★	性質・関係性の把握	行列分解（非負値行列因子分解、特異値分解）をツールを使って実行でき、その結果を正しく解釈できる			
71	13	性質・関係性の把握	★★★	性質・関係性の把握	空間的自己相関の手法を用いて空間的な類似性を数値化できる			
72	14	性質・関係性の把握	★★★	性質・関係性の把握	行列分解（非負値行列因子分解、特異値分解）を、目的に応じてパラメータを最適化して分析できる			
73	1	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			○
74	2	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			
75	3	サンプリング	★★	サンプリング	調査対象の母集団の規模・特性や調査コストに応じて、多段階抽出法や層化抽出法など適切な標本抽出方法を計画できる			
76	4	サンプリング	★★	サンプリング	属性数と水準数が決まれば適切な直交表を選択し実験計画ができる			
77	5	サンプリング	★★	サンプリング	調査に求められる信頼水準・誤差率から必要となるサンプル数を試算できる			○

NO	Sub No	スキルカテゴリ	スキルレベル	サブカテゴリ	チェック項目	BC	DE	必須スキル
78	1	データ加工	★	データ加工	各義尺展の変数をダミー変数に変換できる			○
79	2	データ加工	★	データ加工	標準化とは何かを知っていて、適切に標準化が行える			○
80	3	データ加工	★	データ加工	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・変換などの対応ができる			○
81	4	データ加工	★	データ加工	加工済データに分析上の不具合がないか検証できる			○
82	5	データ加工	★★	データ加工	各義故の分布・欠損率などをふまえて、外れ値・異常値・欠損値の対応を決定できる			○
83	6	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、必要に応じて層的変数のカテゴリライズを設計・実行できる			
84	7	データ加工	★★	データ加工	加工データに不具合がないか自分でテストを設計し、検証できる			
85	8	データ加工	★★★	データ加工	加工データの統計的な前提によって不具合の早期発見ができるとともに、統計的観点で次ステップの解析に耐えうるデータであるか評価できる			
86	1	データ可視化	★	方向性定義	可視化の目的の広がりや概略について説明できる（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）	*		
87	2	データ可視化	★★	方向性定義	特異点を明確にする、データ解析部門以外の方にデータの意味を正しく伝える、現場の作業を支援するといった可視化の役割・方向性を判別できる	*		
88	3	データ可視化	★★★	方向性定義	データ量が膨大で構造が捉えにくい場合や、アウトプットが想像しにくい場合であっても、可視化の役割・方向性を判断できる（ビッグデータ中の要素間の関連性をダイナミックに表現する、細部に入りすぎずに間に対して答えを出すなど）			
89	4	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			○
90	5	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			○
91	6	データ可視化	★★	軸だし	抽出したい意味にふさわしい軸・層化の粒度、順番を考慮して軸のきざみや層化方法を選択できる			○
92	7	データ可視化	★★	軸だし	膨大な属性を持つテーブルから目的に有用な属性を選択できる			
93	8	データ可視化	★★★	軸だし	非構造データから分析の軸になりうる候補を抽出し、付加すべき属性候補を適切に出せる			
94	9	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			
95	10	データ可視化	★★	データ加工	データの持つ分散量の観点で、高次のデータを主成分分析（PCA）などにより1～3次元のデータに変換できる			
96	11	データ可視化	★★	データ加工	統計量を使うことで、データの読み取りたい特徴を効果的に可視化できる			
97	12	データ可視化	★★★	データ加工	非線形（高次の曲線、渦状の分布など）のデータであっても、高次のデータの次元を、次元圧縮（1～3次元のデータに変換）して、特徴（データの縦分散量および各データの位置関係）を損なわずに可視化できる			
98	13	データ可視化	★★★	データ加工	ネットワーク構造、グラフ構造などの表現において、ノードとエッジが稠密すぎて特徴抽出が困難であっても、データの絞り込みや抽象度を上げることで適切に可視化できる	*		
99	14	データ可視化	★★★	データ加工	データ量が膨大（ペタバイト以上）なために、処理しきれず描画できない規模のデータに対しても、適度なデータや情報の抽出（掘引き）、クラスタリングなどにより可視化しうる状態にデータを加工できる	*		
100	15	データ可視化	★★★	データ加工	大規模なデータへのリアルタイムな可視化が求められる場合であっても、特異点の抽出や次元圧縮を通じてデータを圧縮し、リアルタイム表示できる	*		

NO	Sub No	スキルカテゴリ	スキルレベル	サブカテゴリ	チェック項目	BC	DE	必須スキル
101	16	データ可視化	★★★	データ加工	大規模なデータへのリアルタイムな可視化が求められる場合であっても、データの分割転送、復元を通じて可視化できる		*	
102	17	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）			
103	18	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）			○
104	19	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）	*		○
105	20	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）に応じた図表化ができる			○
106	21	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表現するなど）			
107	22	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ラインとしての可視化ができる	*		
108	23	データ可視化	★	表現・実装技法	ポロノイ図の概念と活用方法を説明できる			
109	24	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（平行座標、散布図行列、テーブルレンズ、ヒートマップなど）を適切に可視化できる			○
110	25	データ可視化	★★	表現・実装技法	ネットワーク構造、グラフ構造、階層構造などの統計的な関係性の可視化ができる	*		○
111	26	データ可視化	★★	表現・実装技法	GPSデータなどを平面地図上に重ね合わせた可視化ができる			
112	27	データ可視化	★★	表現・実装技法	挙動・軌跡の可視化ができる（店舗内でのユーザーの回遊やEye trackingなど）			
113	28	データ可視化	★★	表現・実装技法	適切な情報（意味）を押さえた上で、デザイン性を高めるための要件提示ができる			
114	29	データ可視化	★★★	表現・実装技法	人体、標高を持つ地図、球面などの上にデータを重ね合わせた可視化ができる		*	
115	30	データ可視化	★★★	表現・実装技法	地図上で同時に動く数百以上のポイントにおける時間変化を動的に表現できる（多地点での風の動き、飛行物の軌跡など）		*	
116	31	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			○
117	32	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			
118	33	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ことができる（特異点、相違性、傾向性、関連性を見出すなど）			
119	34	データ可視化	★★	意味抽出	統計値（代表値の指標、バラツキの指標、有意性の指標、関係式）を正しく読み、図解式や移動平均線に意味付けできる			
120	35	データ可視化	★★★	意味抽出	分類系の分析において、分布傾向から原因を追究、活用（分類に応じたDM発送による反応率の向上など）、ドリルダウンを計画し主導できる			
121	36	データ可視化	★★★	意味抽出	予測系の分析において、関連性、特異点、変曲点から原因を追究、活用（予測結果に基づく発注管理など）を計画し主導できる			
122	37	データ可視化	★★★	意味抽出	関連系の分析において関連が高い/低い原因、活用（リコメンドなど）、ドリルダウンを計画し主導できる			
123	1	分析プロセス	★	アプローチ設計	スコープ、検討範囲・内容が明快地設定できれば、必要なデータ、分析手法、可視化などを適切に選択できる	*		
124	2	分析プロセス	★★	アプローチ設計	解くべき課題がフレーミングされていれば、必要なデータ、分析手法、可視化などを適切に選択できる	*		○
125	3	分析プロセス	★★★	アプローチ設計	複数の事業や課題にまたがっていても、必要なデータ、分析手法、可視化などを適切に選択し作業手順に落とし込める	*		
126	4	分析プロセス	★★★	アプローチ設計	複数のアプローチの組み合わせでしか解けない課題であっても、その解決までの道筋を設計できる	*		
127	5	分析プロセス	★★	分析価値の判断	分析で解くべき課題が否かを判断できる	*		

NO	Sub No	スキルカテゴリ	スキルレベル	サブカテゴリ	チェック項目	BC	DE	必須スキル
128	1	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	*		
129	2	データの理解・検証	★★	統計情報への正しい理解	積極的に統計情報を収集しているとともに、表現に惑わされず数字を正當に評価できる（原点が0ではないグラフ、不当に誇張されたグラフなど）	*		
130	3	データの理解・検証	★★★	統計情報への正しい理解	数字やデータの検証のために、何と比較するべきかすみやかに把握し、収集・利用できる（業務データや過去に接触した統計情報の想起・活用を含む）	*		
131	4	データの理解・検証	★	データ確認	単独のグラフに対して、集計ミスなどがないかチェックできる	*		○
132	5	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	*		
133	6	データの理解・検証	★★	データ確認	複数のグラフや集計表で構成されているレポートに対して、全体として集計ミスや不整合が起きていないかチェックできる	*		○
134	7	データの理解・検証	★★	データ確認	データ項目やデータの量・質の検証方法を計画・実行し、その結果をもとにその後の分析プロセスを立案・修正できる	*		
135	8	データの理解・検証	★★★	データ確認	多数のグラフ、集計表、外部の統計情報、高度なデータ解析手法を用いた解析結果などを含むレポートに対して、不整合が起きていないか、妥当性の高い論理構造であるかチェックできる	*		
136	9	データの理解・検証	★★★	データ確認	分析に必要なデータを想定し、現在取得可能なデータの量・質で分析に耐えうるか、分析目的が達成可能であるかを判断できる	*		○
137	10	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、駒呑みにはしないことの重要性を理解している	*		
138	11	データの理解・検証	★★	俯瞰・メタ思考	データを俯瞰して、変化をすみやかに察知するとともに、変化が誤差の範囲かどうか判断できる	*		○
139	12	データの理解・検証	★★★	俯瞰・メタ思考	複数のデータを多元的かつ大域的に俯瞰して、大きな動きや本質的な事実を見抜くことができる	*		○
140	13	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	*		○
141	14	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	*		○
142	15	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	*		○
143	16	データの理解・検証	★★	データ理解	生データを眺めて、どのような切り口で集計・比較すればデータの理解や事実の把握につながるかを検討できる	*		○
144	17	データの理解・検証	★★	データ理解	扱ったことのない新たなデータに内容の不明な項目があっても、生データの閲覧や集計を通して何の項目かあたりをつけられる	*		
145	18	データの理解・検証	★★	データ理解	扱っているデータの関連業務の知識と分析目的を踏まえて、どんな説明変数が効きそうか、あたりをつけて洗い出し、構造的に整理できる	*		
146	19	データの理解・検証	★★★	データ理解	データの変化から起きている現象の背景を構造的に推察し、仮説を立て、検証方法を企画実行できる	*		
147	20	データの理解・検証	★★★	データ理解	データを入手する前に、存在するであろうデータとその分布を想定して基礎的な方向性やその結果の想定ができ、それを前提とした解析方法の検討・ラフ設計をすることができる	*		
148	21	データの理解・検証	★★★	データ理解	扱ったことのない新たなデータであっても、ER図やテーブル定義、生データなどを見ることによってデータの発生源や欠損値の意味などのあたりをつけられる	*		
149	22	データの理解・検証	★★	データ粒度	分析目的とデータの量・質を踏まえて、想定されるメッセージと統計的観点から適切な集計単位とサンプリング率を決定できる	*		
150	23	データの理解・検証	★★★	データ粒度	分析目的とデータの量・質に加えて、想定しているメッセージ、深掘りの方向性・可能性、処理負荷、データ処理フローなども総合的に踏まえた最適な集計単位とサンプリング率を決定できる	*		
151	1	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）	*		○

NO	Sub No	スキルカテゴリ	スキルレベル	サブカテゴリ	チェック項目	BC	DE	必須スキル
152	2	意味合いの抽出、洞察	★	洞察	想定に影響されず、分析結果の教鞭を客観的に解釈できる	*		
153	3	意味合いの抽出、洞察	★★	評価	各種の解析手法（主成分分析、クラスター分析、決定木分析など）の結果を解釈し、意味合いを適切に表現・説明できる	*		
154	4	意味合いの抽出、洞察	★★	評価	分析結果が当初の目的を満たしていない場合に、問題を正しく理解し、目的達成に向けて必要な分析手順を追加・変更できる	*		
155	1	機械学習	★	機械学習	機械学習にあたる解析手法（Random Forestなど）の名称を3つ以上知っており、手法の概要を説明できる			
156	2	機械学習	★	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か理解している			
157	3	機械学習	★★	機械学習	「教師あり学習」「教師なし学習」「強化学習」の違いを理解しており、データの内容や学習手法に応じて適切な学習データ、テストデータ、検証データ（チューニング用データ）を作成できる			
158	4	機械学習	★★	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			○
159	5	機械学習	★★	機械学習	回帰予測モデルの検討において、過学習を防止するためL1正則化（Lasso回帰）、L2正則化（Ridge回帰）を適切に適用できる			
160	6	機械学習	★★	機械学習	ROCカーブを用いてモデルの精度を評価できる			
161	7	機械学習	★★	機械学習	混同行列（正誤分布のクロス表）を用いてモデルの精度を評価できる			
162	8	機械学習	★★	機械学習	MSE（Mean Square Error）、Accuracy、Precision、Recall、F値といった評価尺度を理解し、実際の精度評価を行うことができる			
163	9	機械学習	★★	機械学習	サポートベクターマシンによる分析を、統計解析ツールを使って実行でき、その結果を解釈できる			
164	10	機械学習	★★	機械学習	アンサンブル学習（Random Forest、勾配ブースティングマシン、バギング）による分析を、統計解析ツールを使って実行でき、その結果を正しく解釈できる			
165	11	機械学習	★★	機械学習	ニューラルネットワークによる分析を、統計解析ツールを使って実行でき、その結果を正しく解釈できる			
166	12	機械学習	★★★	機械学習	課題やデータ型に応じて、サポートベクターマシンの適切なモデルを選定し、目的に応じてアルゴリズムの調整や設計ができる			
167	13	機械学習	★★★	機械学習	課題やデータ型に応じて、アンサンブル学習（Random Forest、勾配ブースティングマシン、バギング）の適切なモデルを選定し、目的に応じてアルゴリズムの調整や設計ができる			
168	14	機械学習	★★★	機械学習	CNN、RNN/LSTMなどの深層学習（ディープラーニング）の主要方式の特徴を理解し、目的に応じて適切に選定できる ※CNN：Convolutional Neural Network （畳み込みニューラルネットワーク） ※RNN：Recurrent Neural Network （再帰型ニューラルネットワーク） ※LSTM：Long Short-Term Memory （長期短期記憶）			
169	15	機械学習	★★★	機械学習	深層学習（ディープラーニング）の実装において、予想精度を向上するため、層の種類（プール、畳み込み）、層数、ニューロン数、活性化関数、学習回数などをチューニングできる			
170	16	機械学習	★★★	機械学習	ホールドアウト法、交差検証法などを用いて、モデルの汎化能力評価ができる			
171	17	機械学習	★★★	機械学習	過学習を回避する方法を設計・実施できる			
172	18	機械学習	★★★	機械学習	高次元データの取り扱いについて、次元の呪いを考慮し適切に次元削減できる			
173	19	機械学習	★★★	機械学習	機械学習等の最新の論文を理解し、必要とあれば自分で実装し評価できる			

NO	Sub No	スキルカテゴリ	スキルレベル	サブカテゴリ	チェック項目	BC	DE	必須スキル
174	20	機械学習	★★★	機械学習	バイアスとバリエーションの関係を理解し、モデル選定を適切に行える			
175	1	時系列分析	★	時系列分析	時系列データについて説明ができる（時系列グラフ、周期性、移動平均など）			
176	2	時系列分析	★★	時系列分析	時系列データの時点差での相関関係を、系列相関やコレログラムを利用して評価ができる			
177	3	時系列分析	★★	時系列分析	時系列データに対し、ツールを使用して、分析結果の比較を行い、適切なモデルを選択できる（自己回帰モデル、移動平均モデル、ARIMA、SARIMA、指数平滑法など）			
178	4	時系列分析	★★	時系列分析	時系列分析は少なくとも3つの要素の視点をもって行うべきことを理解している（長期トレンド、季節成分、その他周期性など）			
179	5	時系列分析	★★	時系列分析	ピリオドグラムにおいて、FFT（Fast Fourier Transform）など計算量を抑制する方法により、ピリオドグラムの計算ができる			
180	6	時系列分析	★★★	時系列分析	状態空間モデルにおいて、カルマンフィルタを用いて、観測値から欠測値の補間をし、予測モデルを構築できる			
181	7	時系列分析	★★★	時系列分析	非線形・非ガウス型状態空間モデルにおいて、モンテカルロ・フィルタを用いて、複雑な時系列システムの予測モデルを構築できる			
182	1	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			
183	2	言語処理	★★	言語処理	形態素解析や係り受け解析のツールを適切に使い、基本的な文書構造解析を行うことができる			
184	3	言語処理	★★	言語処理	TF-IDFやcos類似度などの基本的なアルゴリズムを使い、単語ベクトルの作成や文書群の類似度計算を行うことができる			
185	4	言語処理	★★★	言語処理	形態素解析・構文解析・固有表現抽出のアルゴリズムを理解し、使いこなせる			
186	5	言語処理	★★★	言語処理	N-gram言語モデルの構築方法と代表的なスムージングアルゴリズムを理解し、使いこなせる			
187	6	言語処理	★★★	言語処理	索引型の全文検索の仕組み（転置インデックス、スコアリング、関連性フィードバック）を理解し、使いこなせる			
188	7	言語処理	★★★	言語処理	Trie、Suffix Arrayなどの代表的な高速文字列検索アルゴリズムを理解し、使いこなせる			
189	8	言語処理	★★★	言語処理	潜在的意味解析（LSA）の仕組みを理解し、使いこなせる			
190	9	言語処理	★★★	言語処理	データの特性に合わせ、適切な言語処理アルゴリズムを選択し、誤り分析、辞書作成などを行い、成果を最大化することができる			
191	10	言語処理	★★★	言語処理	トピックモデル、サポートベクターマシン（SVM）などの文書分類手法を理解し、実行できる			
192	11	言語処理	★★★	言語処理	再帰型ニューラルネットワーク（RNN）、長期短期記憶（LSTM）などを用いたニューラルネットワーク型言語モデルを理解し使いこなせる			
193	12	言語処理	★★★	言語処理	隠れマルコフモデル（HMM）などを用いた系列ラベリング手法を理解し使いこなせる			
194	13	言語処理	★★★	言語処理	スキップグラム（Skip-gram）などの分散表現モデルを理解し使いこなせる			
195	1	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			
196	2	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			
197	3	画像・動画処理	★★	画像処理	画像の処理や解析において、効果的なパターン検出や画像特徴抽出などを既存手法から選ぶことができる		+	
198	4	画像・動画処理	★★	画像処理	画像の処理や解析において、既存のAPI化したクラウドサービスなどを目的に即して、選定・活用することができる			
199	5	画像・動画処理	★★★	画像処理	物体検出・識別などの画像処理手法に関して、適切な論文などの文献を参考に実装し評価できる			

NO	Sub No	スキルカテゴリ	スキルレベル	サブカテゴリ	アビリティ説明	BZ	DE	必須スキル
200	6	画像・動画処理	★★★	画像処理	画像・動画処理を行う環境に合わせて適切な実装・手法を選ぶことができる		*	
201	7	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			
202	8	画像・動画処理	★★★	動画処理	動画の自動解析手法の状況について理解し、適切な専門家のサポートの元で実装を検討できる			
203	1	音声/音楽処理	★	音声/音楽処理	wavやmp3などの代表的な音声フォーマットを知っている			
204	2	音声/音楽処理	★★	音声/音楽処理	音声データから、分析目的にあった波形データの抽出やノイズの除去をすることができる			
205	3	音声/音楽処理	★★	音声/音楽処理	音声認識や本人認証、感情分析などの代表的な音声処理分野について理解し、用いられる分析手法を説明することができる			
206	4	音声/音楽処理	★★★	音声/音楽処理	ケプストラム分析やLPC（線形予測分析）などの代表的な音声信号分析手法を理解し、使いこなすことができる			
207	5	音声/音楽処理	★★★	音声/音楽処理	音声認識や認証・感情分析などの目的に合わせて、パラメータ調整や手法変更、言語モデル・音響モデルなどを差し替え、モデル構築・精度評価をすることができる			
208	1	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			
209	2	パターン発見	★★	パターン発見	アプリオリアルゴリズムのアソシエーション分析において、統計ツールのパラメータをチューニングして出力件数を調整できる			
210	3	パターン発見	★★	パターン発見	レコメンデーション業務の要件から適切な協調フィルタリングロジック（ユーザーベース、アイテムベースなど）の選定を行うことができる			
211	1	グラフィカルモデル	★★	グラフィカルモデル	パス解析において、変数間の因果関係をパス図を用いて説明できる			
212	2	グラフィカルモデル	★★	グラフィカルモデル	ベイジアンネットワーク分析結果から目的事象の事後確率を算出できる			
213	3	グラフィカルモデル	★★	グラフィカルモデル	ツールを用いて共分散構造分析（構造方程式モデリング：SEM）を行い、観測変数・潜在変数の因果関係を説明できる			
214	1	シミュレーション/ データ同化	★★	シミュレーション/ データ同化	離散型・連続型シミュレーションについて、説明できる （モンテカルロ、ヒストリカル、Agentベースなど）			
215	2	シミュレーション/ データ同化	★★	シミュレーション/ データ同化	マルコフ連鎖の特徴を理解し、MCMC（マルコフ連鎖モンテカルロ法）シミュレーションをツールを用いて実装できる			
216	3	シミュレーション/ データ同化	★★★	シミュレーション/ データ同化	データ同化の概念を理解し、実行できる（データを用いてシミュレーション内の不確実性を減少させる計算技法など）			
217	4	シミュレーション/ データ同化	★★★	シミュレーション/ データ同化	シミュレーションにおける問題を理解し、対処を考えることができる （初期条件・境界条件・パラメータの不確実性、データ分布の不均一性、実験計画の最適性など）			
218	5	シミュレーション/ データ同化	★★★	シミュレーション/ データ同化	MCMC（マルコフ連鎖モンテカルロ法）における各種アルゴリズム（メトロポリス・ヘイスティングス法、ギブスサンプラー、ハミルトニアン・モンテカルロ法など）について理解し、活用できる			
219	1	最適化	★★	最適化	線形計画法について、説明することができる			
220	2	最適化	★★	最適化	複数のA/Bテストの統計的結果を踏まえ、デザイン等の最適化を行う手法を回すことができる		*	
221	3	最適化	★★	最適化	一定の制約下で最適解の識別と報酬の最大化がともに求められ、かつ報酬分布が時間経過で変化するような問題に対して、多腕バンディットアルゴリズムを適用・実装できる			
222	4	最適化	★★	最適化	凸関数および、凸計画問題の条件や特徴を説明できる			
223	5	最適化	★★	最適化	連続最適化問題（制約なし）において、使用可能なアルゴリズムを説明することができる（ニュートン法、最急降下法など）			
224	6	最適化	★★	最適化	連続最適化問題（制約あり）において、使用可能なアルゴリズムを説明することができる（ラグランジュ未定乗数法、内点法、逐次2次計画法など）			

NO	Sub No	スキルカテゴリ	スキルレベル	サブカテゴリ	チェック項目	BZ	DE	必須スキル
225	7	最適化	★★	最適化	組み合わせ最適化問題において、代表的な解法の概念を説明することができる（厳密解法（分枝限定法、動的計画法、切除平面法）、近似解法（局所探索、貪欲法など）、メタヒューリスティック解法（遺伝的アルゴリズム、タブーサーチなど））			
226	8	最適化	★★★	最適化	ビジネス課題にあわせて、変数、目的関数、制約を定式化し、線形・非線形を問わず、最適化モデリングができる			
227	9	最適化	★★★	最適化	代表的な最適化問題に関して、モデリングを行い、ソルバーを使い、最適化できる（ナップザック問題、ネットワークフロー問題、巡回路問題など）			
228	10	最適化	★★★	最適化	バッチ勾配降下法、確率的勾配降下法、ミニバッチ勾配降下法の違いを説明でき、勾配降下法の複数のアルゴリズムを、目的に応じて使い分けることができる（Momentum, Adamなど）			

スキルチェックリスト 2017年 改訂版 <データエンジニアリングカ>

▼ 該当のスキル

NO	Sub No	スキルカテゴリ	スキルレベル	サブカテゴリ	チェック項目	BZ	DS	必須スキル
1	1	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			
2	2	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			
3	3	環境構築	★★	システム運用	数千レコードのデータを保持するシステムにおいてデータの重要性や分析要件に則したシステム構築、初期データ投入方法、システム運用の要件定義が行える			○
4	4	環境構築	★★	システム運用	顧客管理など分析システムの運用（異常検知、フェイルオーバー、バックアップ、リカバリー処理、開始・停止処理）の手動書作成や要件定義が行える			
5	5	環境構築	★★	システム運用	数千レコードのデータを保持するシステムのキャパシティ要件（データ容量）と必要処理性能（スループット）を定義できる			
6	6	環境構築	★★★	システム運用	数十億レコードのデータに対してリバランシングなども含めてシステム拡張方法や最適化の要件を整理できる			
7	7	環境構築	★★★	システム運用	扱うデータのデータ規模や機密性、分析要件を理解した上で、オンプレミスで構築するか、クラウド上で構築するかの要件を整理できる			
8	8	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			○
9	9	環境構築	★	システム企画	オープンデータ活用目的でExcelを使った分析システムの要件定義が行える			
10	10	環境構築	★★	システム企画	HTTPを活用したオープンAPIと分析システムのサーバー環境及びデータベースの連携設計ができる			
11	11	環境構築	★★	システム企画	社内分析者向けのRDBMS、NoSQL、ETL、Visualizationなど単一コンポーネントのユーザー利用機能設計が行える			○
12	12	環境構築	★★	システム企画	ソフトウェア開発プロジェクトの管理方法、設計、テスト方法を理解した上で、データ管理・分析システムを要求定義することができる			
13	13	環境構築	★★	システム企画	深層学習（ディープラーニング）の学習を高速化するために、GPU（GPGPU）環境を設計・実装できる			
14	14	環境構築	★★★	システム企画	数十億レコード規模のデータに対し、HadoopやSparkなどを組み合わせた適切なシステム構成を設計できる			
15	15	環境構築	★★★	システム企画	組織を横断する多種多様なデータが混在する環境下でのデータ活用ニーズに対して適切なシステム環境を設計・提案できる			
16	16	環境構築	★★★	システム企画	数十億レコードの内部データ（CRM、webログ、ユーザー購買データ）、外部データ（購入しているデータ、オープンデータ）を理解し、複数のデータソースを統合する要件を整理できる			
17	17	環境構築	★★★	システム企画	サービス上のそれぞれの機能がどのデータに関連があるか把握し、分析機能追加やシステム変更の要件を整理できる			○
18	18	環境構築	★★★	システム企画	KVS、カラム指向、ドキュメント指向などデータ構造の異なる複数のシステムからデータ取得と分析環境への連携が設計できる			
19	19	環境構築	★★★	システム企画	数十億レコードのデータを持つ業務要件やリソース負荷に応じて、データフローや管理機構の統合、またバッチ実行スケジュールの最適化について要件定義が行える			
20	20	環境構築	★★★	アーキテクチャ設計	全体システム化計画及び個別システム化構想・計画を具体化するために、全体最適の観点を持ちながら、対象とするデータ分析システムの開発に必要な要件を整理することができる			
21	21	環境構築	★★★	アーキテクチャ設計	データ活用戦略を正しく理解し、各種業務プロセスについての専門知識とシステムに関する知識を有し、双方を活用して、適切な要求定義が行える	*		
22	1	データ収集	★	クライアント技術	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			○
23	2	データ収集	★★	クライアント技術	対象プラットフォーム（iOS、Android、HEMSなど）におけるデータ取得の仕様（精度など）を理解しており、システム要件を満たせるか判断できる			

NO	Sub No	スキルカテゴリ	スキルレベル	サブカテゴリ	アフェクト項目	UZ	DS	必須スキル
24	3	データ収集	★★	クライアント技術	目的に選したログ取得項目を、対象プラットフォーム (iOS、Android、HEMSなど) で取得可能なデータを用いて設計できる			
25	4	データ収集	★★	クライアント技術	対象プラットフォームにおけるバッテリー消費や通信速度などを含めたシステム要件を作成できる			
26	5	データ収集	★★	クライアント技術	Webクローラーの仕組みを理解し使いこなせる			
27	6	データ収集	★★★	クライアント技術	取得するデータが増えることを想定し、拡張性を考慮したクライアントアプリケーション (ロガーなど) を設計できる			
28	7	データ収集	★★★	クライアント技術	プラットフォームの違いを吸収し、統一的にデータを取得するプログラムを実装できる			
29	8	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			
30	9	データ収集	★★	通信技術	データ収集対象の要件に応じて、MQTT (Message Queue Telemetry Transport) によるパブリッシュ/サブスクライブ型の通信を検討・実装できる			
31	10	データ収集	★★	通信技術	ネットワークプロトコルや暗号化などの通信技術を用い、通信のボトルネックと可用性 (継続的に通信が成立していること) を考慮した上で、必要な通信機能を実装できる			
32	11	データ収集	★★★	通信技術	データ通信において、機能・性能の問題に対し根本原因を特定できるだけでなく、必要に応じて新規技術の適用を検討できる			
33	12	データ収集	★★	データ収集	既存のサービスやアプリケーションに対して、有効な分析をするためのログ出力の追加仕様を整理することができる	*	*	
34	13	データ収集	★★★	データ収集	入手可能なデータに加え、分析結果の品質・効果を向上させる新たなデータ入手方法を提案できる (IoTでの新設センサーの種類・配置場所・データ入手間隔など)	*	*	○
35	14	データ収集	★	データ統合	同種のデータを統合するシステムを設計できる			○
36	15	データ収集	★★★	データ統合	異種フォーマットが混在するデータを統合するシステムを設計できる			
37	16	データ収集	★★★	データ統合	システム分析・業務分析をもとに、必要なデータフロー管理やジョブ管理ツールを選定・評価できる			
38	1	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) なのか非構造化データ (雑多なテキスト、音声、画像、動画など) なのか判断できる			○
39	2	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			○
40	3	データ構造	★	要件定義	業務で使用するシステムのデータのライフサイクル (いつ、どんなデータが発生し、いつまで保持されているのかなど) を把握して、論理モデルを作成できる			○
41	4	データ構造	★★	要件定義	データ保持ルール (データアクセス、性能、保持期間、セキュリティなど) に基づき、データベース・DWHの運用ルールを定義できる			
42	5	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			
43	6	データ構造	★★	テーブル定義	ビジネスプロセスを理解・整理して、データフロー図、論理データモデル、ER図、テーブル定義書を作成できる	*		○
44	7	データ構造	★★	テーブル定義	業務特性や基幹システムの特徴をもとに、検索で頻発に使用するデータのキー (顧客IDなど) を想定し、インデックスを作成・設定できる			
45	8	データ構造	★★	テーブル定義	データ集計を高速化またはSQLを単純化するため、スタースキーマ、スノーflakeスキーマなどを用いたデータモデルを設計できる			
46	9	データ構造	★★	テーブル設計	稼働中の複数のシステム間で発生するデータ項目の差異を、変換テーブルを活用して、埋めることができる			
47	10	データ構造	★★	テーブル設計	DWHに入れる元データ (基幹DBのデータなど) のキーに変更があった場合に、サロゲートキーやナビゲーションブリッジテーブルを用いて対応できる			

NO	Sub No	スキルカテゴリ	スキルレベル	サブカテゴリ	アペック項目	UZ	DS	必須スキル
48	11	データ構造	★★★	テーブル設計	非正規化テーブルや一時テーブルなどを作成し、アプリケーションの処理速度を高速化できる			
49	1	データ蓄積	★	DWH	DWHアプライアンス（Oracle Exadata、IBM Pure Data/Netezza、Teradataなど）に接続し、複数テーブルを結合したデータを抽出できる			
50	2	データ蓄積	★★	DWH	DWHアプライアンス（Oracle Exadata、IBM Pure Data/Netezza、Teradataなど）の機能と特徴を理解し、適切な管理対象データを選定できる			
51	3	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			○
52	4	データ蓄積	★	分散技術	NoSQLデータストア（HBase、Cassandra、Amazon DynamoDB、Cloudant、Azure DocumentDBなど）にAPIを介してアクセスし、新規データを登録できる			
53	5	データ蓄積	★★	分散技術	分散処理環境のディストリビューションを導入できる（Hortonworks、CDHなど）			
54	6	データ蓄積	★★	分散技術	Hadoopの得意な点、苦手な点を理解し、Hadoopにて管理すべきデータを選定できる			
55	7	データ蓄積	★★	分散技術	KVSの特性（集計・ソートが苦手、データの一貫性保証など）を理解し、KVSがデータストア要件を満たすかを判断できる			
56	8	データ蓄積	★★★	分散技術	分散クラスタ構成が可能なRDBMS製品（Oracle RAC、DB2 Pure Scaleなど）を用いてスケールアウト可能なオンプレミス構成を設計できる			
57	9	データ蓄積	★★★	分散技術	Hadoop・Sparkの分散アーキテクチャを理解し、大容量データ処理のパフォーマンスチューニングができる			
58	10	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、Bluemix Cloud Object Storageなど）に接続しデータを格納できる			
59	11	データ蓄積	★★	クラウド	クラウド上のDWHサービス（Amazon Redshift、Google BigQuery、IBM dashDBなど）にデータをロードし公開できる			
60	12	データ蓄積	★★★	クラウド	クラウド上のデータストアサービスが機能面・非機能面で対象業務に合致するかの評価を行い、採用可否を判断できる			
61	13	データ蓄積	★★★	キャッシュ技術	基盤設計において、どこにシステム要素にmemcachedなどのキャッシュ機能を採用すると処理が高速化されるか判断できる			
62	14	データ蓄積	★★★	リアルタイムデータ分析	リアルタイムに入力されるストリームデータから指定条件のイベントを即時に抽出する複合イベント処理（CEP）を実現するサーバー環境・構成を設計できる			
63	15	データ蓄積	★★★	リアルタイムデータ分析	リアルタイムに連続して入力されるストリームデータの加工・集計処理を行うにあたり、Storm等のリアルタイム分散フレームワーク適用の有効性を判断できる			
64	16	データ蓄積	★★★	グリッドコンピューティング	分散処理のフレームワーク（Spark、Tezなど）を用いてアプリケーションの計算処理を複数サーバーに分散させる並列処理システムを設計できる			
65	17	データ蓄積	★★★	新規技術	データストアの技術動向に注目し、リレーショナルDBだけでなく、グラフDB・時系列DBなどの新規技術の検証・評価ができる			
66	1	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）			○
67	2	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をするなど）			
68	3	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る。また、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			○
69	4	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、自己結合ができる。また、UNION処理ができる			○
70	5	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、または既定値に変換できる			○
71	6	データ加工	★★	クレンジング処理	フラットファイルやバイナリファイルに対するデータロードの前処理（クレンジング操作、索引処理やバイナリ処理）ができる			

NO	Sub No	スキルカテゴリ	スキルレベル	サブカテゴリ	アエック項目	UZ	DS	必須スキル
72	7	データ加工	★★	クレンジング処理	縦形補間など、複数のレコードを考慮したクレンジング処理ができる			
73	8	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、規定されたリストと照合して変換する、都道府県名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			○
74	9	データ加工	★★★	マッピング処理	データ定義や実際の観測データの状況をもとに、名寄せ処理を設計・実装できる			
75	10	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			○
76	11	データ加工	★	集計処理	数十万レコードのデータを集計して、合計や最大値、最小値、レコード数を算出できる			○
77	12	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			○
78	13	データ加工	★★	変換・演算処理	数十万レコードのデータに対して、カラム名型のデータに変換できる			
79	1	データ共有	★	データ出力	加工・分析処理結果をCSV、XML、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			○
80	2	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			○
81	3	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			
82	4	データ共有	★★	データ展開	利用者の要件に合致したレポート (図、表) を、PDFやPostScriptなどの印刷用フォーマットで出力する変換機能を設計できる			
83	5	データ共有	★★	データ展開	BIツール用のサーバーを構築し、単一データソースのレポート用DBを設計・構築できる			
84	6	データ共有	★★	データ展開	データストア上のデータをメールやメッセージ (Webサービスなど) を用いてプッシュ配信するシステムのサーバー・ネットワーク・ソフトウェアの構成を設計できる			
85	7	データ共有	★★	データ展開	Webアプリケーションの実装において、WebSocketを用いてクライアント側にリアルタイムにデータ提供できる機能を設計できる			
86	8	データ共有	★★	データ展開	RSS、RDFや業界標準フォーマットなど要件に合致したデータ形式・配信形式で、情報提供するシステムのインターフェースを設計できる			
87	9	データ共有	★★	データ展開	Web API (REST) やWebサービス (SOAP) などを用いて、必要なデータを提供するシステムの公開インターフェースを設計できる			
88	10	データ共有	★	データ連携	FTPサーバー、ファイル共有サーバー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			○
89	11	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			
90	12	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			
91	13	データ共有	★★	データ連携	連携対象システムの仕様に合わせて、ETLツールを用いたデータ変換、ファイル転送処理を実装できる			
92	14	データ共有	★★★	データ連携	ESB・EAIなどのデータ連携基盤を活用してシステム間のデータ連携 (データ配信・交換) を行うインターフェースを設計できる			
93	1	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			○
94	2	プログラミング	★	データインタフェース	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			
95	3	プログラミング	★★	アルゴリズム	非効率なループ処理の性能向上などのために、計算量やメモリを意識したプログラム実装ができる			
96	4	プログラミング	★★★	アルゴリズム	取組の論文に発表された数式処理や機械学習ロジックをプログラム実装できる		+	

NO	Sub No	スキルカテゴリ	スキルレベル	サブカテゴリ	アビリティ説明	UZ	DS	必須スキル
97	5	プログラミング	★★	拡張プログラミング	データ型を意識したプログラム設計・実装ができる（C言語での性能・誤差を意識したデータ型の実装、Pythonでのエラー時のデータ型を意識したデバッグなど）			
98	6	プログラミング	★★	拡張プログラミング	異なるタイプの複数の処理を効率よく行うために、スクリプトを用いたプログラムを設計・実装できる（パイプライン処理のluigiなど）			
99	7	プログラミング	★★	拡張プログラミング	GPU（GPGPU）を有効に活用できるライブラリを選択し、利用できる			
100	8	プログラミング	★★★	拡張プログラミング	GPU（GPGPU）環境において、演算速度を最適化するライブラリを有効に選択し、利用できる		*	
101	9	プログラミング	★★★	拡張プログラミング	対象プラットフォーム（iOS、Android、HEMSなど）におけるデータ取得の業界標準を理解しており、今後の技術動向や規制についてのリスクを提示できる			
102	10	プログラミング	★★	データ規模	SNSから抽出した非構造化データを、適切な言語やライブラリを選んでプログラム実装できる			
103	11	プログラミング	★★★	データ規模	単一サーバーの物理メモリを超える複数のデータソースを組み合わせたデータ処理において、分散処理アーキテクチャやデータのインメモリ処理の特性を意識してプログラム設計ができる			
104	12	プログラミング	★★	分析プログラム	分析プログラムのロジックと処理手順を理解した上で正しい分析結果を出力しているか検証ができる		*	○
105	13	プログラミング	★★★	リアルタイム処理	ストリーミング処理や複合イベント処理（CEP）などを設計し、適切な言語やライブラリを選んでプログラム実装できる			
106	14	プログラミング	*	SQL	SQLで簡単なSELECT文を記述・実行できる（副問合せを含まない、2テーブル程度の結合と簡単なWHERE条件を含むSELECT文）			○
107	15	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、各種JOINの使い分け、集計関数とGROUP BY、CASE文を使用した縦横変換、副問合せやEXISTSの活用など）			○
108	16	プログラミング	★★★	SQL	RDBにおける分析関数の構文と挙動を理解し、分析関数を用いて複雑な副問合せや自己結合を解消できる			
109	17	プログラミング	★★★	SQL	N：Nの結合や完全外部結合の危険性（計算量の増大、結果の不完全性）、暗黙の型変換の危険性（インデックス不使用、小数点以下の切り捨てなど）を考慮したSQLを記述できる			
110	18	プログラミング	★★★	SQL	記述したSQLの実行計画の検証と判断ができ、SQLの修正やインデックス作成により、処理時間を大幅に改善するようなパフォーマンスチューニングができる			
111	19	プログラミング	★★	Pig/HiveQL/Spark SQL	HiveQLやSpark SQLを記述して、パーティションが切られているデータを適切に処理できる			
112	20	プログラミング	★★	Pig/HiveQL/Spark SQL	Pigを記述して列数・内容の異なる行が混在しているデータセットやネスト構造を持つデータセットを処理できる			
113	21	プログラミング	★★★	Pig/HiveQL/Spark SQL	HiveQL、Spark SQL、またはPigで使用するためのUDFが実装できる			
114	22	プログラミング	★★	分散処理	Scala言語を用いて、分散処理環境（Sparkなど）におけるロジックを設計・実装できる			
115	1	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			○
116	2	ITセキュリティ	★★	プライバシー	データ匿名化の方法を理解し、匿名化方法（ハッシュ化、マスキング、k-匿名化など）に応じた加工処理を設計できる			○
117	3	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	*	*	○
118	4	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って設定できる			
119	5	ITセキュリティ	★★	攻撃と防御手法	DoS攻撃、不正アクセス、マルウェア感染や内部不正などのセキュリティインシデントが発覚した場合に既存のルールに基づき対応できる			
120	6	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設計できる			

NO	Sub No	スキルカテゴリ	スキルレベル	サブカテゴリ	チェック項目	UZ	DS	必須スキル
121	7	ITセキュリティ	★★	攻撃と防御手法	SQLインジェクションやバッファオーバーフロー攻撃の概要を理解し、防止する対策を判断できる			
122	8	ITセキュリティ	★★★	攻撃と防御手法	なりすまし、改ざん、盗聴などのセキュリティ侵害を防御するための対策を特定できる			
123	9	ITセキュリティ	★★★★	攻撃と防御手法	侵入検知システム（IDS）やファイアウォールなどを用いて、外部からの不正アクセスを検知、防御する環境を設計できる			
124	10	ITセキュリティ	★★★★	攻撃と防御手法	個別の案件ごとに、依頼元との契約約款、依頼元がデータをどのように保持し利用するかに応じて、適切な匿名化の手法を選択し適用できる	*		
125	11	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			○
126	12	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			
127	13	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			
128	14	ITセキュリティ	★★	暗号化技術	SSHやSSL/TLSなどのセキュアプロトコルの概要と必要性を説明できる			
129	15	ITセキュリティ	★★	暗号化技術	DES、AES、RC4を用いたKerberos認証が使われる事例と仕組みを説明できる			

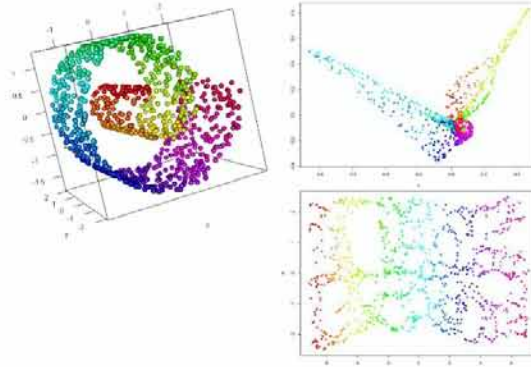
【添付別紙】

スキルチェックリスト：データサイエンティスト

「データ可視化」に関するイメージ共有のための参考資料（例示）

項目NO

10 データの持つ分散量の観点で、高次のデータを主成分分析（PCA）などにより1～3次元のデータに変換できる

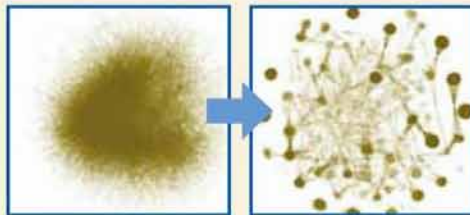


引用元：株式会社ALBERT WEBサイト
http://www.albert2005.co.jp/analyst_blog/?p=1214

13 ネットワーク構造、グラフ構造などの表現において、ノードとエッジが増えすぎて特徴抽出が困難であっても、データの絡り込みや抽象度を上げることで適切に可視化できる

図2 企業間取引データをクラスタ化して把握しやすくした例

	研磨前	研磨後
頂点数	3,282	3,282
枝数	35,168	73,132
フリーク数	32,953	343



引用元：（出版物）大学共同利用機関法人 情報・システム研究機構 国立情報学研究所 発行
「国立情報学研究所ニュース [NII Today] 第66号 平成26年12月」 6ページ
http://www.nii.ac.jp/userdata/results/pr_data/NII_Today/66/all.pdf

【添付別紙】

スキルチェックリスト：データサイエンスカ

「データ可視化」に関するイメージ共有のための参考資料（例示）

項目NO

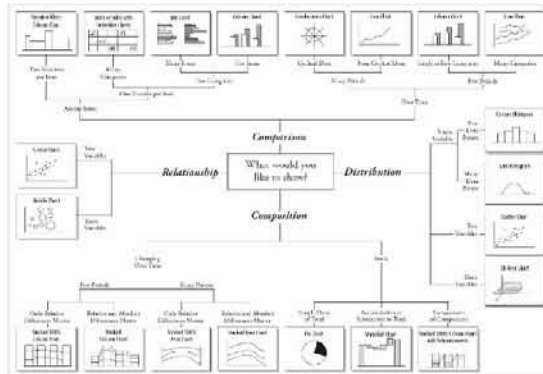
19 強調表現がもたらす効果と、明らかに不適切な強調表現を理解している（計量データに対しては位置やサイズ表現が色表現よりも効果的など）

表現例	量を表現する	順序を表現する	分類を表現する	関係を表現する
位置 	✓	✓	✓	✓
長さ 	✓	✓		
明度/彩度・色 		✓		✓

引用元：「solutions-2」WEBサイト

<http://www.solutions2.be/Books.aspx> をもとに独自に修正加筆

20 1～3次元の比較において目的（比較、構成、分布、変化など）に応じた図表化ができる



引用元：Amit Agarwal 「Digital Inspiration」WEBサイト

<http://www.labnol.org/software/find-right-chart-type-for-your-data/6523/>

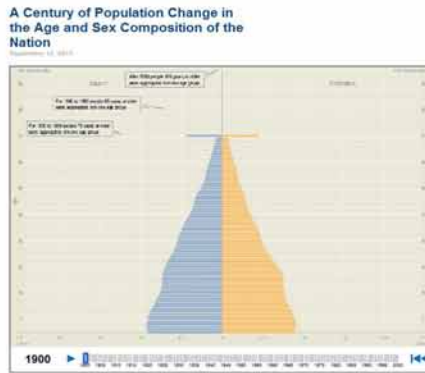
【添付別紙】

スキルチェックリスト：データサイエンスカ

「データ可視化」に関するイメージ共有のための参考資料（例示）

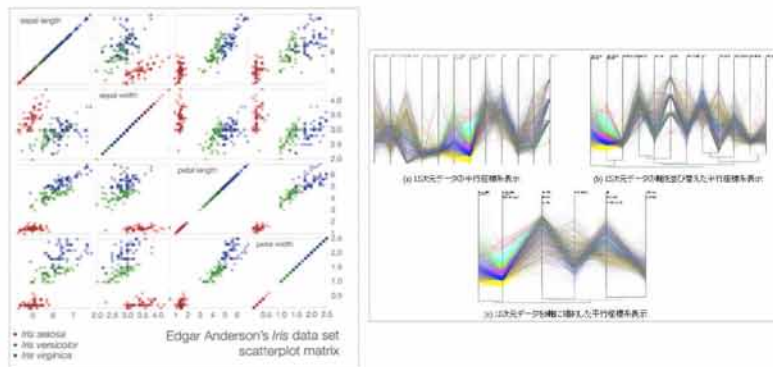
項目NO

21 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表現するなど）



引用元：アメリカ国税調査局 WEBサイト
<http://www.census.gov/dataviz/visualizations/055/>

24 1～3次元の図表を拡張した多変量の比較（平行座標、散布図行列、テーブルレンズ、ヒートマップなど）を適切に可視化できる



引用元：【左図】「Data-Driven Documents」WEBサイト <http://mbostock.github.io/d3/talk/20111116/iris-splom.html>
【右図】「グラフスペクトル解析を用いた軸縮約可能平行座標系」WEBサイト
<http://user.keio.ac.jp/~yun/projects/cpc/index-j.html>

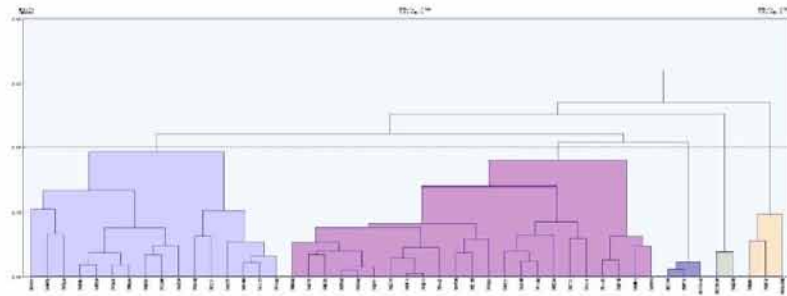
【添付別紙】

スキルチェックリスト：データサイエンサ

「データ可視化」に関するイメージ共有のための参考資料（例示）

項目NO

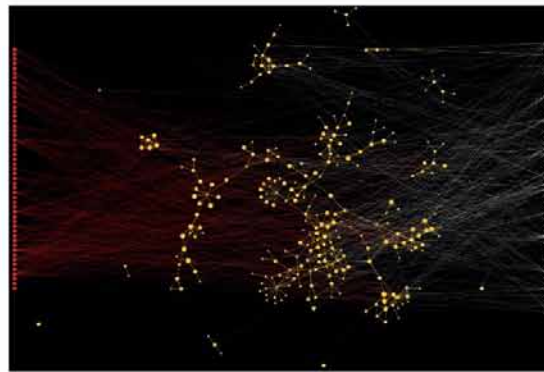
25 ネットワーク構造、グラフ構造、階層構造などの統計的な関係性の可視化ができる



引用元：株式会社ALBERT WEBサイト

http://www.albert2005.co.jp/technology/images/tech_mining_img106.jpg

25 ネットワーク構造、グラフ構造、階層構造などの統計的な関係性の可視化ができる



赤・白点：Wine

黄点：Cheese


引用元：「Cytoscape.js」WEBサイト（Wine & cheese）

<http://cytoscape.github.io/cytoscape.js/demos/cde4db55e581d10405f5/>

【添付別紙】
 スキルチェックリスト：データサイエンス力
 「データ可視化」に関するイメージ共有のための参考資料（例示）

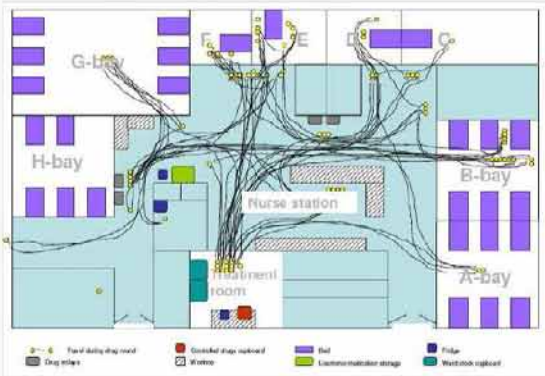
項目NO

26 GPSデータなどを平面地図上に重ね合わせた可視化ができる



引用元：「A WORLD OF TWEETS」WEBサイト
<http://aworldoftweets.frogdesign.com/>

27 挙動・軌跡の可視化ができる（店舗内でのユーザーの回遊やEye trackingなど）



引用元：「Imperial College London」WEBサイト
<https://www1.imperial.ac.uk/cpsq/ms/m/>

【添付別紙】

スキルチェックリスト：データサイエンسカ

「データ可視化」に関するイメージ共有のための参考資料（例示）

項目NO

30 地図上で同時に動く数百以上のポイントにおける時間変化を動的に表現できる（多地点での風の動き、飛行物の軌跡など）



引用元： Kaspersky Lab 「CYBERTHREAT REAL-TIME MAP」 WEBサイト
<https://cybermap.kaspersky.com/>

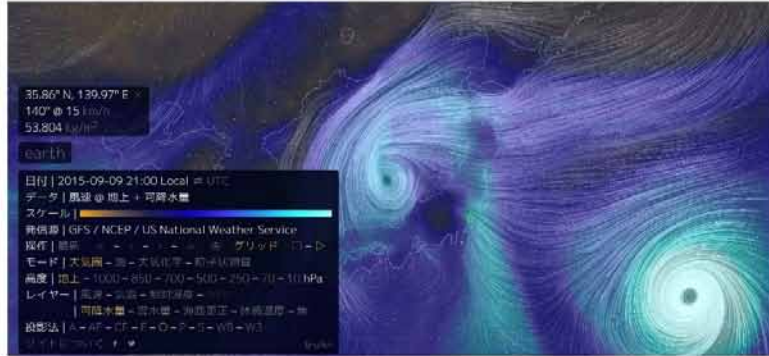
【添付別紙】

スキルチェックリスト：データサイエンسカ

「データ可視化」に関するイメージ共有のための参考資料（例示）

項目NO

30 地図上で同時に動く数百以上のポイントにおける時間変化を動的に表現できる（多地点での風の動き、飛行物の軌跡など）



引用元：Cameron Beccario 「earth」WEBサイト

http://earth.nullschool.net/jp/#2015/09/09/1200Z/wind/surface/level/overlay=total_precipitable_water/orthographic=-224.70,38.05,3000/grid=on

(参考) データサイエンス領域のタスクリスト <IPA ITSS+ (プラス) 2017年4月7日>

<https://www.ipa.go.jp/jica/na/mplus.html>

<IPA: ホームページより抜粋>

「データサイエンス領域」は、企業等の業務における大量データを分析し、その分析結果を活用するための適切なタスクとそのために習得しておくべきスキルを取りまとめた。

この領域は、IPAがこれまで公開してきたITSSには含まれていない、新規の領域です。

タスクは、IPAと「一般社団法人データリエンジニアリング協会（委員長：安部和人、ヤフー株式会社 CSO）」の協業で作成しました。スキルは関係者が公開している「スキルチェックリスト」を参照します。

タスク分類	タスク小分類	タスク小分類	No	評価項目
分析プロジェクトの立ち上げと計画 およびの管理設計	分析プロジェクトの立ち上げと計画 およびの管理設計	前倒案件の明確化	1	分析プロジェクト（データサイエンスを含む）、課題解決を行う一連の取組）のステークホルダーを明らかにする
		目標の明確化	2	分析プロジェクトの目標や期待効果を明確に目的とゴールを明らかにし、ステークホルダー間で共有する
			3	分析プロジェクトの目標を設定する
		推進体制設計	4	分析プロジェクトの目標に評価方法を具体化し、定量的な成功基準を設定するとともに、成功基準の特定時期・測定者も決定する
			5	分析プロジェクトの複雑さに応じ、実施体制と役割分担を、メンバーの様々なデータ分析スキル、チームの過去の経験やパフォーマンスを考慮し、決定する
			6	分析プロジェクトの実施計画を作成する
		計画の承認	7	分析プロジェクトに必要なコストと分析プロジェクトの実施によって得られる利益（コスト削減効果を含む）を分析プロジェクトの実現性について評価・検討する
			8	分析プロジェクトの実現性について評価・検討する
			9	分析プロジェクトの実現計画について、ステークホルダーに説明し、必要に応じて調整を行うことで合意を形成する
			10	分析プロジェクトの実現計画について、ステークホルダーの承認を得る
集約整理	11	分析プロジェクトに必要なハードウェア環境を設計・整備する		
	12	分析プロジェクトに必要な環境構築を設計・整備する		
	13	分析プロジェクトに必要なソフトウェア環境を設計・整備する		
	14	不必要な情報の漏洩がないよう、利用可能な情報源を事前確認し、事前承認を得る		
目的達成後の業務 移行	15	データ分析結果を利用・活用する対象業務のプロセス再構築設計する		
	16	取組データ別にデータの活用用途や用途のカテゴリと連携・アクセス方法をステークホルダー間で設定する		
	17	対象業務の運用体制や運用方法を決定する		
	18	対象業務の移行計画を確定し、モニタリング方針・方法やモニタリング時のKPIを決定する		
データの作成と収集	データ分析設計	19	データの収集方法、加工方法、分割、結合、重複・保存方法などの分析プロセスを決定する	
		20	対象業務に必要となるデータの種類の対象業務の目的に合ったデータ分析手法及びセキュリティ要件（正確性・信頼性、厳密性、保守性など）を検討する	
		21	セキュリティの脆弱性、評価手法（脆弱性診断、脆弱性評価、クラウドリスク評価など）、検証方法（クロスレビュー、シミュレーション）、モニタリングなどによる継続的改善のためのデータ分析手法を検討する	
	データ収集	22	対象業務の目的を踏まえて利用・活用可能なデータまたは収集方法（WEBデータの収集、システムからのデータの出力・加工、外部データの連携・連携等）を決定する	
23	検討したデータが現実的に利用・活用可能なかどうかを検証・検討し、利用可能なデータを収集する			
高度化データ加工	データ加工	24	集約や欠損、外れ値、異常値など、除外または補正するべきデータのクレンジング方針を検討し実行する	
		25	分析目的に照らし、必要十分なサンプリングデータを集約計画法などを用いて作成する	
		26	分析目的に照らし、必要な補綴及び分布等を踏まえ、データを適正に整形する (例) データ結合、データ変換、転送	
		27	利用範囲別のデータマート等の要件を明確化する	
分析用データ準備	入力データと目的データとの データの連携	28	利用範囲別のデータマート等を設計・整備する	
		29	必要に応じて利用範囲別のシステムやインターフェースを準備する	
		30	目的達成、及びその必要要件（期待しない学習など）を定義し、目的達成データと期待達成の元となる入力データを作成する（目的達成については、必要の検証がなければ十分時点や期間も考慮する）	
データ解析	データ設計・抽出	31	セキュリティを確保するために十分なデータ（学習データ）の検証に必要データ（モデルを構築・チューニングするための検証データや、検証確認を行うためのテストデータなど）に分類するとともに、モニタリング手法に適切な必要なデータ追加を行う（データの正規化、データのシャッフルなど）	
		32	データ設計・抽出の設計をする（検索条件・抽出条件、集約回数などを決定する）	
		33	データ設計・抽出も実行する	
		評価・検証性の担保	34	収集したデータの集約設計（平均値、中央値、分散値、標準偏差、分散など）を確認し、全体のバラツキや傾向、異常値などを把握する
35	収集したデータを異なるレイヤーで集約し、データが本来持つべき意味や変動目標の観点から適切感がないかどうか			
36	収集したデータに対する再帰的分析（クロス集計、可視化など）を実施し、再帰的な2次元間の関係性を確認する			
37	収集したデータ以上の多変量データに対する分析（平行集計プロット、ネットワーク分析等）を実施し、多変量におけるデータ間の関係性を確認する			
検証	データ前、中、後の関係性（例：因果関係、ボトルネックなど）があるかを事前に検証し、検証方法（手法前、手法中、手法後）を決定する	38	データ前、中、後の関係性（例：因果関係、ボトルネックなど）があるかを事前に検証し、検証方法（手法前、手法中、手法後）を決定する	
		39	異常検知などにおいて、母集団の平均の推定、分散や標準偏差の平均値の推定などの有意性を検証する（t検定、F検定、カイニ検定など）	
		40	品質管理などにおいて、コンパリングを行い、抽出されたサンプルから不具合などの検出を行い、評価する	
		41	AI/AIシステムの開発により得られた結果に対し、回帰分析がない場合、検証を行い、有意性を判断する（カイニ検定）	
予測・判別	分析目的や用途に 予測・判別	42	分析目的や用途に予測・判別を行うための必要となる説明変数データや特徴量を作成する (手法前: p値による選択、正規化による除外、同値、特徴学習による自動作成など)	
		43	(数値予測) 売上予測、顧客予測、売上率予測など数値の予測モデルを構築する (手法前: 因果関係分析、クロスアイツク関係、ニューラルネットワークなど)	
		44	(2値の判別・予測) 与信可否判別や迷惑メール判別、顧客属性など、二つの状態に対する判別・予測モデルを構築する	
		45	(クラス判別・予測) 新規顧客のセグメントや営業の営業行動パターンなど、新たに発生するデータを分類するための判別・予測モデルを構築する (手法前: 決定木、ニューラルネットワークなど)	
46	(時系列予測) 株価や商品売込など、過去の時系列データに基づいて将来の値を予測・変動予測モデルを構築する (手法前: 自己回帰モデル、移動平均モデル、ARIMA、SARIMAなど)			

業務への組み込みと評価	分析評価	96	「仮説の検証」段階に打ち立てた関係性や仮説などの仮説と分析結果を照らし合わせ、活動目的に対する有効性の観点から、その妥当性を判断する
		97	（仮説の検証）データ分析結果から得られた関係性や仮説などの仮説を再検証し、またその仮説は活動目的に有効なものか判断する
	業務への組み込み	98	分析結果も業務で活用するためのソリューション開発を行う（小規模な試行レポートでのゴール検証や大規模ビッグデータシステムへの組み込みなど）
		99	下の検証したKPIに基づき、データ分析結果を活用する業務が当初の想定通りに実施されているかどうかをモニタリング
	組み込み結果の評価	100	当初の目的と照らし、データ分析結果を活用する業務の成果を評価する
		101	業務の成果を踏まえて、必要に応じてデータ処理手法や分析手法等について再検討を行う
	データ分析結果活用による業務改善	102	データ分析結果を活用して、業務上の課題点や課題を明らかにする
		103	明らかになった業務上の課題の件数に対して対応策を検討する
		104	検証された仮説について、業務への反映と改善を図る
	ビジネス上の効果の把握とさらなる	105	データ分析結果の活用による業務上の成果を必要に応じて定期的に報告・フィードバックする
106		ステークホルダーからの意見や要望を収集し、さらなる改善に活かす	

4 教育プログラムの開発

4.1 プログラムの全体骨格

図 18 には、提案する教育プログラムの柱（実務、分野、データ、解析）とそれぞれの主たる教育項目を示す。提案プログラムでは、それぞれの柱を独立に教育するだけでなく、それぞれ相互依存的な形式で教育することで社会科学型データサイエンティストを効果的に育成する。

「実務」では、ビジネスに内在する課題発見に必要な法制度や社会・経済システムに関する知識を教育する。特に、異種・多次元・大規模データを扱うデータサイエンティストに必要な法制度（個人情報保護法や知的財産法等）を学ぶことにより、個人情報を的確に処理したり、開発した技術やビジネスシステムを資産化したりできるような能力の育成を目指す。「分野」では、経営戦略、マーケティング、消費者行動等の社会科学分野に属する各領域の基本的な知識を、実務課題に即した形式で教育する。「データ」では、社会科学分野にあるデータのタイプ、形式等の理解を促すとともに、それらデータの基本的な扱い方を教育する。「解析」では、社会科学分野に内在するデータを解析するために必要となる解析技術（統計モデル、機械学習、人工知能、データマイニング等）の基礎を教育する。ただし、いずれの柱もそれだけで広く、深い分野であるため、全てを短期で修得可能な「履修証明プログラム」として展開することは困難である。

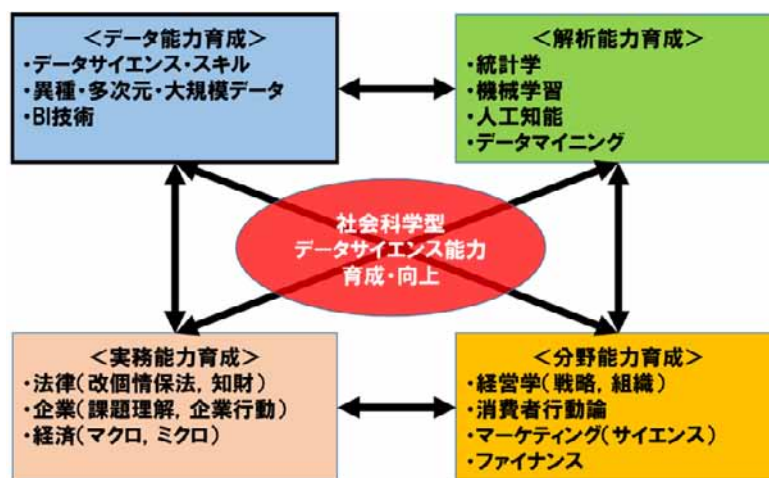


図 18 教育プログラムの骨格

4.2 養成したい人材像

既に紹介したように、社会科学の基礎となるビジネス分野において、異種・多次元・大規模データを単に処理できるだけの人材では不十分である。社会は、ビジネスに内在する課題を適切に把握し、分野の基礎理論でビジネス現象に解釈を与え、それに確認・検証するためのデータ処理および解析を実現できる総合的な能力を有する人材を求めている。総合的な能力を有するデータサイエンティストは、すなわち社会科学分野における課題発見から課題解決、実践までを一気通貫で履行しうる人材である。換言すれば、「マネジメントができるデータサイエンティスト」や「データや解析を高度活用できるマネジメント人材」等がそのイメージになる。

図 19 には、上記を総括的に整理し、本教育プログラムで育成したい人材像を概念的に示した。図 18 に示した個々の柱ごとの教育を実施することで、①実務力（課題の把握、関連法の理解）、②分野力（現象を説明できる知識）、③データ力（活用可能なデータは何か、どう処理すればよいか）、④解析力（データを高度に解析するには）、といった能力をそれぞれ高めることが可能である。さらに、個々の柱を相互依存的に教育することで、本プログラムの学習成果として、①～④の個々の能力の向上に加え、それら能力の総合的活用能力も身につく。すなわち、本教育プログラムを実現できれば、上述した人材を社会に還元できると考えている。

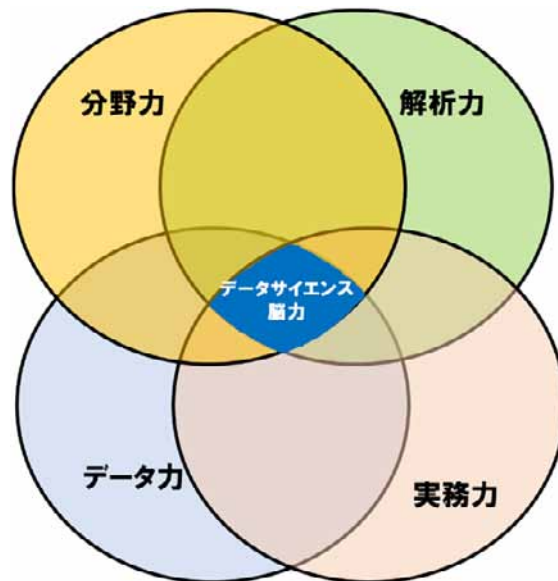


図 19 養成したい人材像

4.3 データ関連業務に従事する社会人を対象とした事前調査結果

4.3.1 調査概要

検討の教育プログラムの方向性が社会的ニーズと合致しているかと検討内容に抜けがないかを確認することを目的とし、データ関連業務に何等かの形で従事している、表 1 に示す 39 社 43 人を対象とし、事前調査を実施した。調査項目は、下記の囲みの通りである。

- ・ 「自社内でデータ処理関連業務は多いかどうか」、「自社内でデータ処理業務を実施する重要度は高いか」、「自社内の経営戦略立案にデータ解析は生かされているか」、「自社内でデータ処理関連業務をする人材の評価はどうか」、「自社内でデータ処理関連人材は足りているか」、「自社内に大規模データが存在するか」
- ・ 「業務上、下記項目を知ることの必要性はどの程度か」
 - 「システム・ソフトウェア開発」、「データの前処理技術」、「機械学習」、「人工知能 (AI)」、「確率・統計」、「統計モデル」、「エージェント・ベース・シミュレーション」、「テキストマイニング」、「データマイニング」、「分野知識 (経営、マーケティング、会計、ファイナンス、経済等)」、「法律的知識 (個人情報保護法案、知財関連)」
- ・ 「自身でデータ処理を実施することが多いか少ないか」および「フリーコメント」

表 1 調査対象者の概要

業種	社数	人数
エンタテインメント	1	1
金融	3	3
広告代理業	4	4
情報産業	6	8
シンクタンク	9	10
製造業(医療機器)	1	1
製造業(音響)	1	1
製造業(自動車)	2	3
製造業(製薬)	1	1
製造業(日雑)	1	1
大学職員	1	1
データベンダー	1	1
農業関連	1	1
不動産業	1	1
マスコミ(IT系)	1	1
マスコミ(新聞)	2	2
マスコミ(テレビ)	1	1
小売業	2	2
合計	39	43

4.3.2 調査結果

① データ処理関連業務を取り囲む状況

[集計結果 (図 20～図 25) の総括 1]

- ・ データ処理関連業務 (多く, 重要度は高い)
- ・ 経営戦略にデータ解析を生かしている傾向
- ・ データサイエンス人材が著しく不足している
- ・ 大規模データが企業には存在している

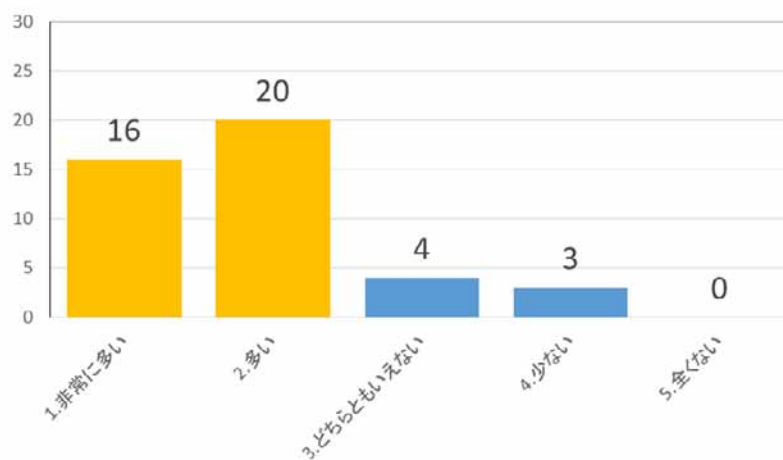


図 20 自社内でデータ処理関連業務は多いかどうか

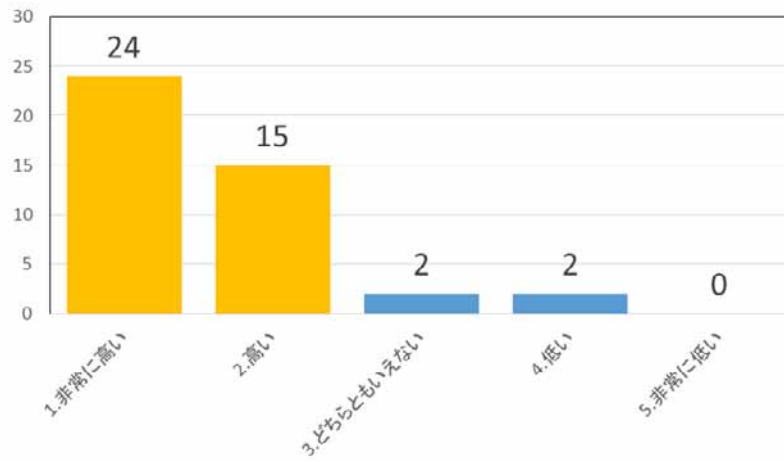


図 21 自社内でデータ処理業務を実施する重要度は高いか

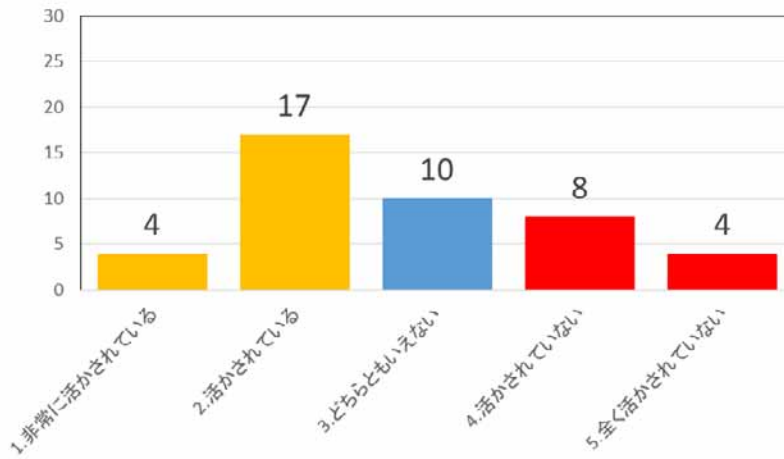


図 22 自社内の経営戦略立案にデータ解析は生かされているか

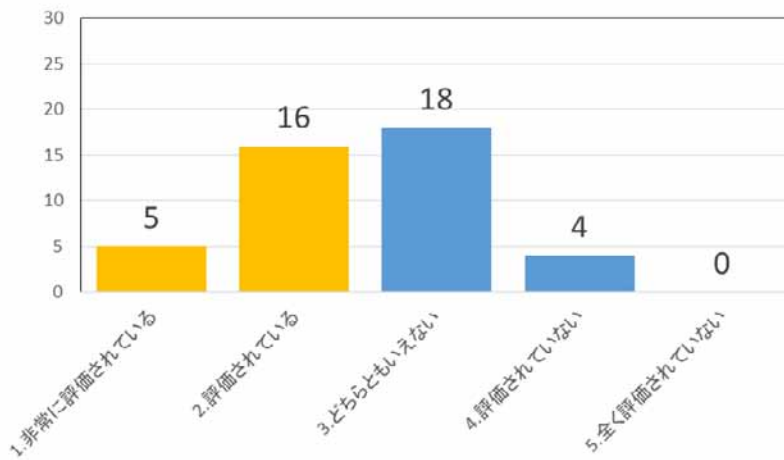


図 23 自社内でデータ処理関連業務をする人材の評価はどうか

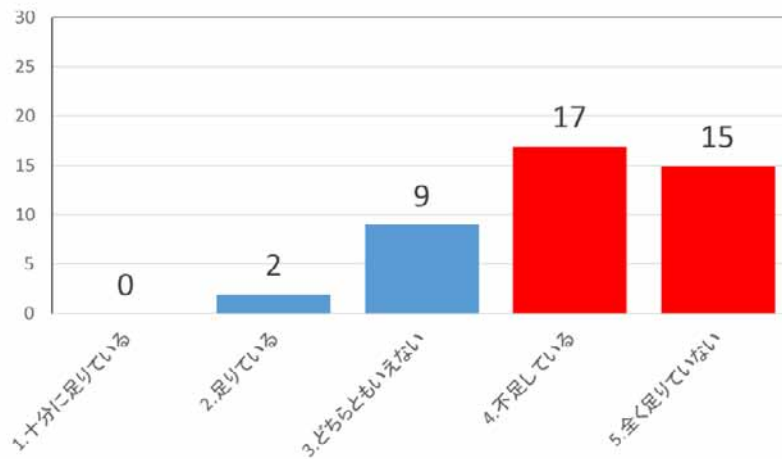


図 24 自社内でデータ処理関連人材は足りているか

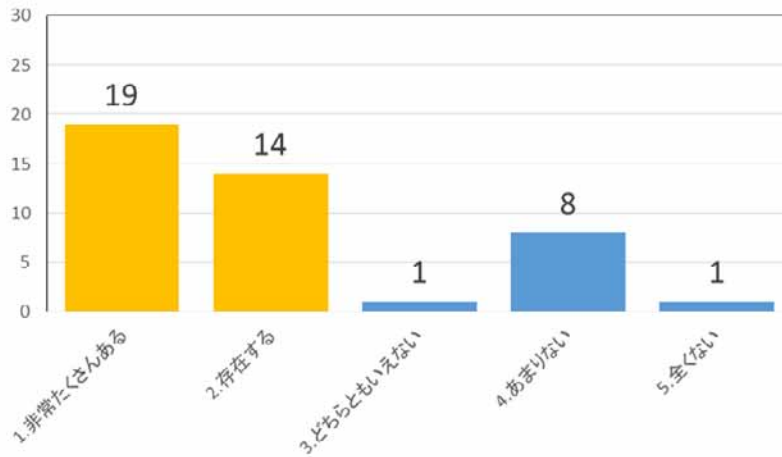


図 25 自社内に大規模データが存在するか

② データ関連業務を遂行する上での科目を知ることの必要性

[集計結果 (図 26～図 36) の総括 2]

- ・ システム・ソフトウェア開発，データの事前処理技術，機械学習，人工知能，確率・統計，統計モデル→必要性が高いという認識
- ・ 特に確率・統計に関しては非常に必要とされている
- ・ テキストマイニング，データマイニング→必要性が高いという認識
- ・ 分野知識，法律的知識→非常に必要とされている

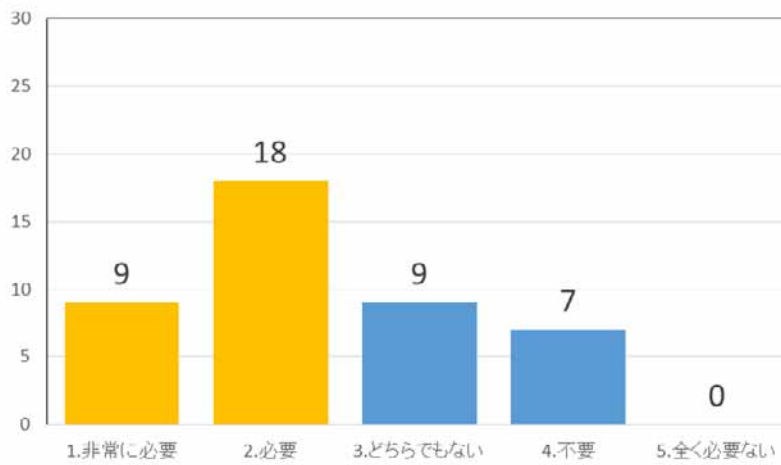


図 26 システム・ソフトウェア開発

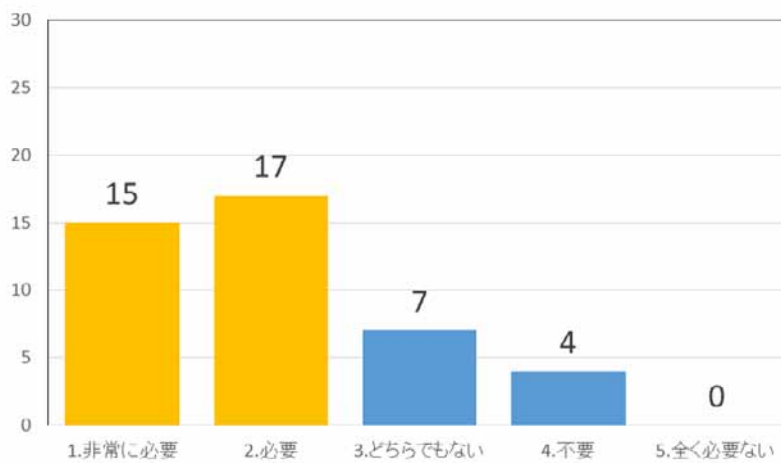


図 27 データの前処理技術

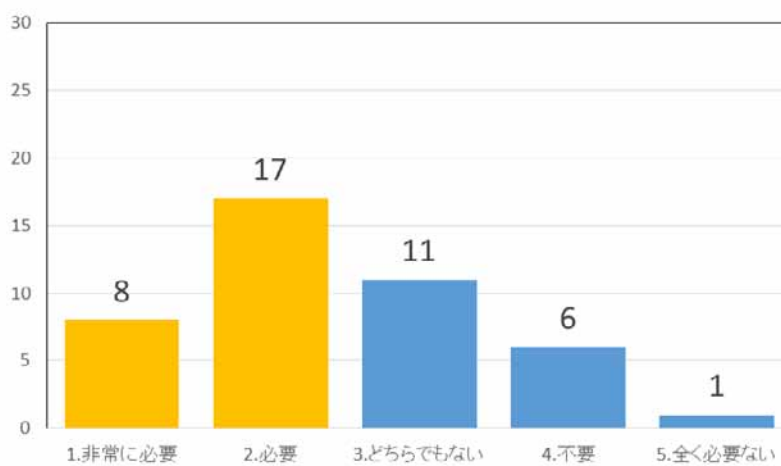


図 28 機械学習

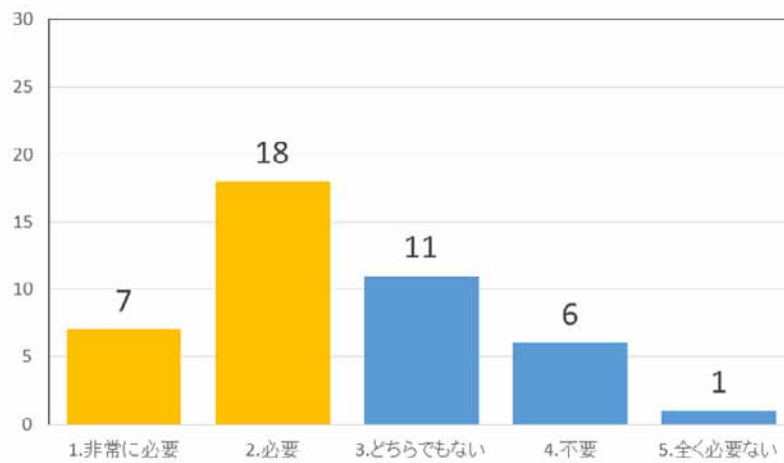


図 29 人工知能 (AI)

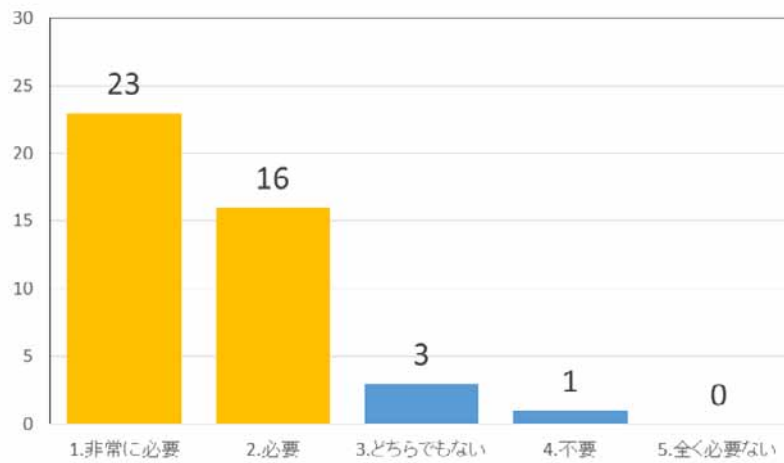


図 30 確率・統計

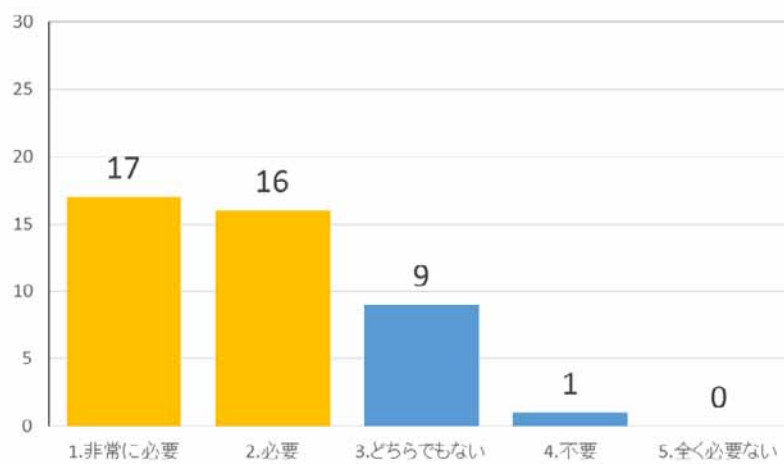


図 31 統計モデル

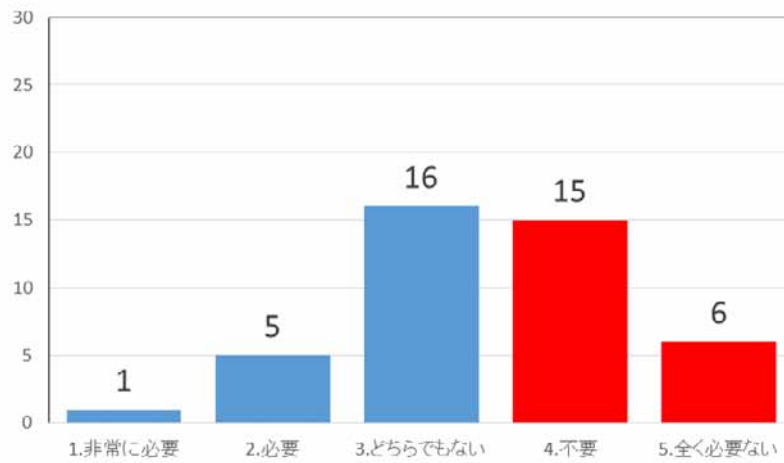


図 32 エージェント・ベース・シミュレーション

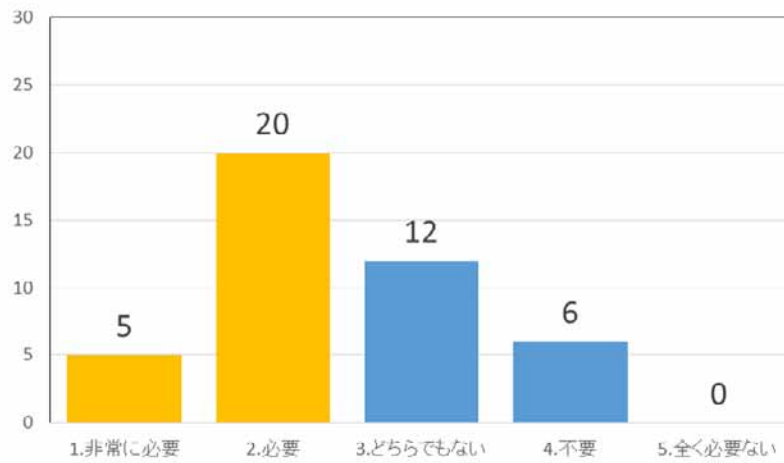


図 33 テキストマイニング

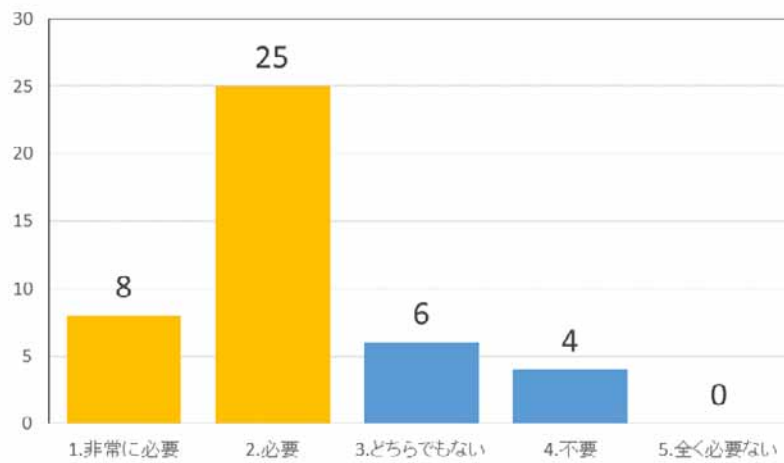


図 34 データマイニング

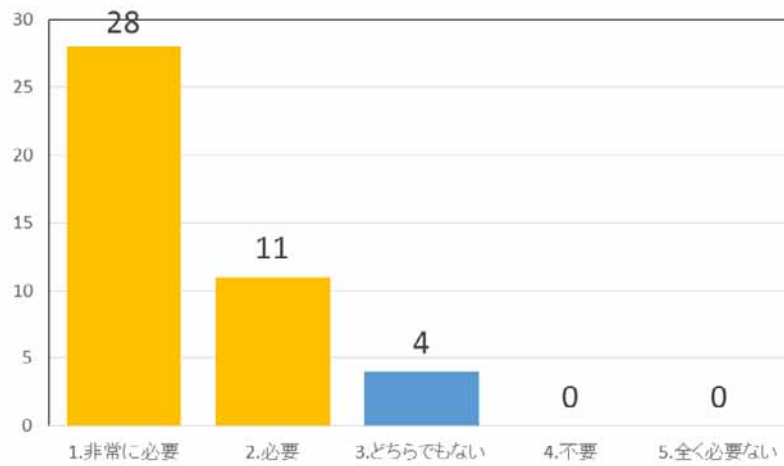


図 35 分野知識（経営，マーケティング，会計，ファイナンス，経済等）

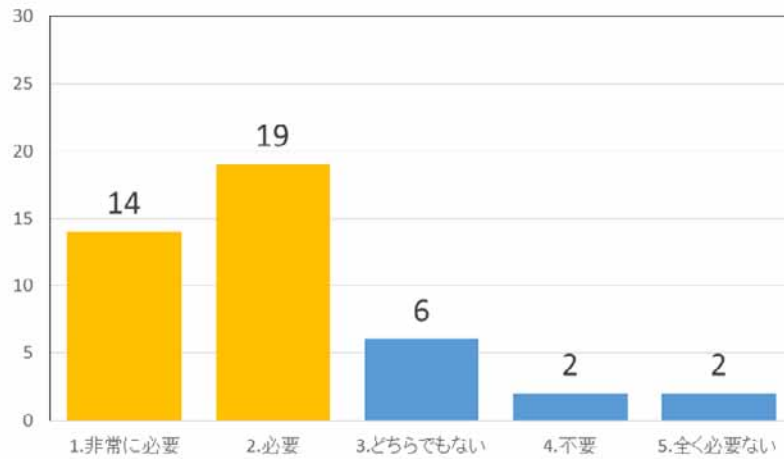


図 36 法的知識（個人情報保護法案，知財関連）

③ フリーコメント

表 2 フリーコメント

業種等	回答(問:企業内で大規模データ等を活かす際の課題や問題点・データサイエンス教育に期待すること)
インターネット関連／外食・レストラン・フードサービス／旅行／エステ・理容・美容	<ul style="list-style-type: none"> ・ 何を課題解決するかの問題設定 ・ 新人の育成計画
コンサルティング	<ul style="list-style-type: none"> ・ クライアント企業とのデータ授受のセキュリティーが甘い反面、複雑になりすぎるとデータ活用促進の妨げになることもある。 ・ 企業で利用するデータ分析ソフトが SAS など、個人で利用するには高価な場合、そもそもの利用できるユーザが限定的となり広がり難い。
製造業(自動車)	<ul style="list-style-type: none"> ・ 部署単位で管理されているデータを統合管理する際のデータクレンジングおよび DB プラットフォーム構築の重要性が理解されていない課題があります。
インターネットサービス業	<ul style="list-style-type: none"> ・ データを用いて説明できることと、データによってはほとんど説明できないことの区別が難しいときがある ・ データを用いて説明しきったり、モデルを無理やりつくろうとする時がある。 ・ よってデータの利活用が誤った方向に行くことがある ・ どの職種もデータサイエンスの初期教育は企業として実施すべきと考える
シンクタンク	<ul style="list-style-type: none"> ・ 仮にそのときに最適なシステムを構築したとしてもそれを使いこなしたり、環境変化に応じて適切にアップデートしていける人材がいない、あるいはいなくなる。
IT	<ul style="list-style-type: none"> ・ 自社ではあまり問題を感じたことがないが、お客様でよくある例として部署毎にサイロになっているため部門や事業所を超えたデータのやり取りが難しく、活用が進まないことがある。
ソフトウェア	<p>弊社においては、上記のようなデータサイエンスの知識の必要性は、アサインされるプロジェクトや顧客案件に大きく依存する。また、社内では、予算管理、売り上げ管理、等の Controlling のデータを扱うことは多いが、高度な統計的分析を戦略的に活用する機会(触れる社員)は少ないと感じる。一方で、どのような業務に携わる場合でも、確率・統計の基本的な考え方は非常に重要であり、今後のデータサイエンス教育の根幹として期待したいところである。なお、機械学習や AI に関しては、世の中の流れとして一通り抑えておくことは有用と考えるが、背後にある数理の部分をしっかり学ぶか、その特徴を理解した上でマネタイズできるようなビジネスモデルを立案することができる人材が求められると考えており、いわゆる API を呼んで機械学習プログラムを作成できるようなエンジニアは寧ろ教育に力を入れなくても今後増加し、コモディティ化すると想定する。</p>

<p>広告代理業</p>	<p>大規模データのシステム(データベースやデータウェア)を構築できる IT 人材と大規模データ処理や分析 (SQL など前処理と R, Python など統計分析)に習熟した人材が、ここ1~2年で急速に必要なになってきている。また、これらを含むビジネス全体を俯瞰し、プロデュースできる人材も求められている。</p>
<p>テレビ</p>	<p><課題></p> <ul style="list-style-type: none"> ・ データサイエンスへの正しい理解 <ul style="list-style-type: none"> → データサイエンスに先行する企業の事例が聞こえてくるようになり経営者含めて必要性は理解しているものの中身を理解せずに「魔法の箱」として捉えられてしまいトップダウンで「人工知能を使って新しいものを・・・」というケースが少なくない。実務レベルも知見を持った人間がおらずベンダー頼みになってしまいがち。 ・ 大規模データを取り扱うにあたって求められる知識領域の広さ <ul style="list-style-type: none"> → ・業務理解と業務システムのデータ構造に関する知識 ・ KPI 設計に関する知識 ・ クラウド利用に伴うクラウドプラットフォーム(AWS/GCP/Azure など) ・ 大規模データに対応するための分散コンピューティング(Hadoop/spark など) ・ リアルタイム処理が求められつつあり、そのジョブ設計とミドルウェア(kafka など) ・ データの前処理技術(ETL)及びデータセット作成・集計(SQL など) ・ 経営層・担当者向けのダッシュボード及び BI ツール ・ 大規模データの高度利活用 <ul style="list-style-type: none"> → 上述のとおり、大規模データを取り扱うには広範な領域の知識・人材が必要となる。これらは売り上げ管理や権利料算定など業務上に必須な領域となることが多く、この領域で手一杯となったり、外部依存してブラックボックス化してしまうことで統計モデルによる知見の抽出や機械学習による予測など高度利活用にまで手が回らない。 ・ 現場のデータリテラシーの低さ <ul style="list-style-type: none"> → 経験上、平均と合計以上の統計量は理解できない。分散は無理です。相関は言葉は知っているが正しく理解していない。また、変数が 3 次元以上になっても理解できない。以前、RFM 分析を行って優良顧客層などクラスターを作成してレポートしたが 3 次元になるとほとんど理解されなかった。逆に、アソシエーション分析で一緒に見られるコンテンツというレポートを出した際は実経験とリンクした結果となったせいか思った以上に理解された。
<p>コンサルティング</p>	<ul style="list-style-type: none"> ・ 分野知識×データサイエンスの組み合わせが重要と思われるが、その育成方法が体系化されていない。現時点では OJT で鍛えるしかない。

新聞社	<ul style="list-style-type: none"> ・ そもそも、経営陣(役員)クラスが、データによる課題の分析、評価に対して理解がない。 ・ 身近にデータ分析を取り入れている部署がない。 ・ データサイエンスを社会人になってから学ぶのは非常にコストと時間がかかり、かつ個人の適正に課題がある一方、気軽に外注することも難しいことが懸念。
金融	課題・問題点:データの寡占化, 教育への期待:データ分析の適用可能範囲と限界, 統計的リテラシーなど
情報サービス業	技術力は高いが、マーケティングや経営学への知識・関心が少ない人材が多いと感じます。データ分析を企業の戦略施策に活かす場合に上記は必須なので、技術面に加えてこれらについても対応していただくと実務の観点からは非常にありがたいです。
メディア・サービス業	複数の業務にまたがるので、多面的な教育、及びその中でもフォーカスした人材育成が必要(例えば、データサイエンティストを ML エンジニアと、マーケティング系サイエンティストに分けるなど)
製造業(自動車)	<p>データサイエンスの実践と内製化に必須となる統計, IT, 現場知識の 3 つのスキルを一定以上有するスタッフ, マネージャ, 上級管理職が不足している。「ビッグデータ」, 「アナリティクス」, 「AI」などのキーワードだけが先行しているため、同分野の業務を受注する外部業者は乱立しているが、社外・国外へのデータ漏洩, モデルやノウハウの流出, ブラックボックス化による品質やコストの不透明さが大変懸念される。そもそも統計学やデータサイエンスの発展は、現実・現場・現物における当面の課題に取り組んだ歴史ととともにあり、これからの日本においても学術と産業の距離を抜本的に縮めることを第一の戦略とすることが、国内産業の競争優位性を確保するために極めて重要となる。具体的には、データサイエンスの研究者は、たまたま入手が容易であったデータや、自身が実証したい理論やモデルに適合するデータを選択し研究対象とするのではなく、潜在性の高いデータと最も困難な課題を抱える企業や組織に出向き、現場と一体となって問題の解決に挑み、その結果として独自の理論やモデルを構築するというアプローチをもっと取るべきであり、それが実現できる環境を整備する必要がある。また、教育においても現実のデータを企業などから入手し、具体的な洞察を得たり課題を解決する実践的なスタイルを目指さなければ、学生はその後戦力となるデータサイエンティストには成り得ず、結局企業内での再教育と選別過程を余儀なくされるため、社会的な非効率を免れない。</p>

<p>広告代理業</p>	<p>とにかく、長時間労働は悪だとされるなか、頼り手は、「ロボット」しかいません。その場合は、実務上使えるロボットを現場で使う事が非常に重要です。ロボットを使いこなすためには最低限の知識と、なによりも、実際の現場でのヘビーな環境のなかでの経験が重要です。これから、ロボットやAIに定型的労働を移管するにあたって、GSSMの役割は超！重要だと思います。生産性向上の為のコースがあってもよいくらい、日本の企業では人手不足と労働時間の減少はまったなしです。実務でロボットを活躍させることができる為の簡易的な研修などを期待します！！</p>
<p>IT 関連</p>	<ul style="list-style-type: none"> ・ 専門知識をもった人材の不足
<p>IT・通信関連</p>	<p>弊社では、データ活用(可視化・分析)を推進する社内横断的組織を昨年立ち上げ、もろもろの課題を解決しながら取り組みを推進しているところです。社内から様々なデータ分析に関する相談を受けており、できれば外部でデータサイエンスに関する専門的な知識を補完できる機関との共同研究等を試みたいと思っています。</p>
<p>製造業(自動車)</p>	<ul style="list-style-type: none"> ・ IoT により車両データは社内に蓄積されつつありますが、そのデータを解析し解釈する人材が不足しております。 ・ 日本ではお客様情報をデータベース化し管理しておりますが、販売マーケティングへの貢献度は低いです。 ・ 海外の販売地域で顧客情報の何をどのように管理し、マーケティングに生かしているかが情報共有されておられません。 ・ 今後、ますます取得できるデータ量は多くなると思いますが、データを適せつに分析し解釈した上で、開発やマーケティングに活かせる人材を育成するという意味でデータサイエンス教育に期待しております。
<p>マーケティング</p>	<ul style="list-style-type: none"> ・ 自社にデータサイエンティストがいない(コンサル会社や研究所にデータ分析を頼んでも、自社の課題や保有するデータの性質やとれるデータの範囲などを完全に理解して最適なデータ分析や DMP 等の設計を行うには限界があるため、内製化が必要だと思う。) ・ ブリッジサイエンティストの存在(経営層など、全くデータ分析の知識がない意思決定層とデータサイエンティストの間をつなぐ役割が必要だと思う。また、データ分析の結果を踏まえて、PDCA を回す現場との間にも同様の役割が必要だと思う。) ・ 不要なビックデータの蓄積(必要最低限のデータ分析でこと足りるものを、わざわざビックデータに広げて分析する昨今のトレンドは非効率だと思う。) ・ オープンソース化の遅れ(自社データ行きすぎた囲い込みが、業界やサービス全体のパフォーマンスを落としていると思う。例えば、患者のカルテデータなどは、オープンにすることで、最適な医療行為が受けられるなど。)

生命保険業	<ul style="list-style-type: none"> ・ ガバナンスが厳しく、データ抽出に時間がかかる ・ 社内でデータサイエンス教育が施されていない ・ データ分析をする人が業界知識や業務等がない場合、効果的な分析や結果が得られない（机上の分析となり、活かしかねないことがある）
広告会社	<p>業務では商品やイベントの感想をアンケートでとることが多いのですが、集計は度数や度数比にとどまっていたり、自由回答もられつにとどまっていたりするものが現実です。これらを感覚による分析ではなく、科学的に説得ある分析ができればビジネス上とても有効であると思います。大規模データにこだわることなく、“小規模データ”についても教育を受ける機会を提供されることをご検討いただきたく思います。</p>
小売業	<p>大量データを処理する仕組みがなく、処理自体に工数にとられるケースも多いように感じます。結果、スピード感が失われ、データ活用の実用性がなくなってしまう。AI や機械学習といった言葉が独り歩きしている感は否めず、ビジネスにおけるデータ活用のあり方、言葉の定義、所謂ビジネス書の一部にあるような通り一遍ではない、実践的なデータサイエンス教育というものがあれば、個人的にも是非受けて見たく、チームメンバーにも勧めたいと思います。</p>
医療	<ul style="list-style-type: none"> ・ 大規模なデータ自体があっても、そもそも活用できるようなデータの取り方が出来ていない ・ 企業のマネジメントに、データサイエンスの重要性への理解が不足している ・ 教育によってデータサイエンス人材が育成されても、データサイエンス人材として採用し評価・処遇できる態勢等が大半の企業で整っていないのでは？
マーケティングリサーチ	<p>デジタル化の風潮にクライアント側もついていけない現状の中、弊社が提供できることを模索中。</p>
マーケティング	<ul style="list-style-type: none"> ・ 組織上の課題 <ul style="list-style-type: none"> － 部門組織を超えた社内包括的なデータマネジメント体制の確立が難しい点 － 経営・マーケティング課題と、データマネジメント上の技術的課題の双方を理解し、課題解決まで進められる人材の不足 ・ 経営・マーケティング上の課題 <ul style="list-style-type: none"> － データマネジメント業務の業務量の方が多く、経営・マーケティング課題解決のためのデータ分析業務に十分な時間・リソースを割けない － スピードが求められる経営・マーケティング上の意思決定における、タイムリーな Data availability の問題 ・ グローバル企業の課題 <ul style="list-style-type: none"> － 各国地域においてデータ環境・データ量・成熟度がまちまちであり、データマネジメントが進んでいる欧米地域のデータと、日本や中国、アジア諸国などとのデータ統合・活用における困難性

研究・コンサルティング	データマイニングであっても、一定の仮説構築力が必要であると思います。エンジニアではなく、サイエンティストであるならば、データの読み方、解釈、その応用方法を考える力も、分析力同様に求められると思います。
化学、日用品	統計解析を行う知識が一部の人に偏っており、それ以外の人にはその重要性が認識され、業務に活用されているとは言い難い。お金を支払えば、調査会社の方がデータの統計解析を請け負ってくれる面もあり(実務上は、その結果の解釈の方が重要)、必要になる頻度もそれほど高くないので、統計解析を一から学ぼうという意欲はそれほど高くない。
大学職員	<ul style="list-style-type: none"> 基礎的な統計知識が社員全員にないとデータ活用の意義や意味が理解できない
シンクタンク	統計改革推進会議をはじめ、行政においてもEBPMの必要性が喧伝されていますが、民間企業でも戦略立案上、エビデンスを求める声はかつてなく高まっています。一方で、それを支えるデータサイエンティスト等の人材には限りがあり、一部には充実した人材がそろっていると思われる企業はあるものの、多くの企業では、そもそも人材がいないか、他の業務に追われてデータの活用には手が回っていないのが現状ではないかと思われます。滋賀大学など学部設立に動いたところもありますが、文理を問わず、大学教育のなかでも、統計に関する講座は必修化するなど、統計人材の育成が進むことを期待しています。 ※企業内に統計人材を育てるカリキュラムをもつところはほとんどなく、外部の研修として期待できる先に心当りのある人事担当者も少ないはずです。
IT系メディア(マスコミ)	<ul style="list-style-type: none"> データの重要性は認識しつつも、意思決定層との知識GAPがあり、従事する社員に対し、正しい評価ができていない。 データを整形、分析には高度なスキルが必要であるのに、意思決定層や周囲はその真価を評価できず、所謂社内下請け、の様な存在になりがち→データやインサイトは出せて当たり前、判断するのは経営側、むしろ経営側が判断しやすいデータをもって来ない場合に対し評価をしまっている現状。
農業関連、申請担当部署 (Regulatory Affairs)	データ解析そのものよりも、データの解釈と結論の導き方について意見が分かれることが多く、サイエンスベースでのデータの理解が課題であるように感じています。特に、有意差の有無だけで状況を判断してしまい、その差がどれだけの影響を及ぼすのか(またはnegligibleなのか)への議論へと進まないことが多いです。
保険業	データ分析を行う人材の育成が課題です。現在は、部署ごとにIBMやデータ分析会社の講座に必要な応じて受講させています。ただ、当会ではデータ分析をおこなう人材を組織的に育成して体制や計画が備わっているわけではないので、データ分析をおこなう部署の上司の方針で、どこまで企業の資源をそこに割り当てるかが左右されてしまいます。

大学生向けウェブサイトの運営・営業	大学生向けウェブサイトの運営・営業をしています。閲覧データを広告などマーケティングにデータを活用できそうだが、まだPV数が少なく、データの取得を含めそうしたことを積極的にはしていない。ただ、それは周囲に専門家がいないからかもしれません。データを活用すれば、状況が改善するのはわかっているが、専門家不足など人的な制約から、できない面があります。
小売業	<ul style="list-style-type: none"> 最低限の統計に関する知識の教育が必要 文系など数学的バックグラウンドがない人材でも短期間で習得するにはどうしたらよいか、最低限の方法があるか
不動産業	<ul style="list-style-type: none"> 分析する際に使用するデータの確からしさ、粒度
研究開発の業務改善、戦略展開	<ul style="list-style-type: none"> データの解釈が今後の教育課題だと思う データをどうやったら使えるのかどうかの判別や、使えるようにするまでの作業に関する教育 データがあるから使う場合と、目的に対してデータを探す・作る、どちらの能力も必要だと思う データサイエンティストも周辺知識を持つことでチームとしての目標達成に貢献出来るのではないかと思う

4.4 開発したプログラムの概要

4.1 項～4.3 項に示した内容を踏まえ、本委託事業では「マーケティング分野のデータサイエンティスト育成」を対象とし、図 37 に示す領域および科目でプログラムを構成した。個々の科目の詳細は、5.1 項で後述する。今回の委託事業では、時間的、施設の制約により基本的には構成したプログラムを講義形式で試行する。

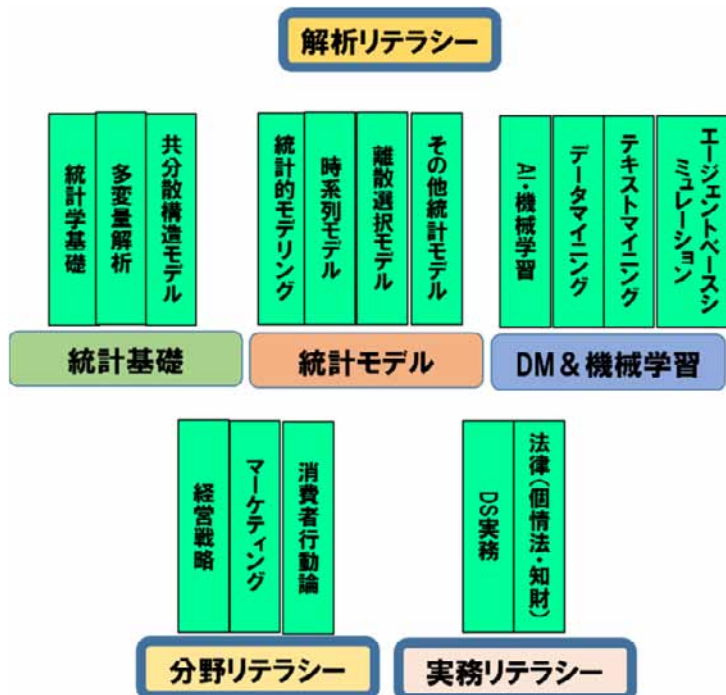


図 37 開発教育プログラムの構造