

# データサイエンティストのための スキルチェックリスト/ タスクリスト概説

Skill Checklist & Task list Overview



一般社団法人  
データサイエンティスト協会  
スキル定義委員会

**IPA**

独立行政法人  
情報処理推進機構

ITSS + (データサイエンス領域)



## はじめに — Preface

世の中は、産業革命以来の歴史的な局面を迎えようとしています。AI (Artificial Intelligence: 人工知能) が人間のトップ棋士に勝利したニュースはまだ記憶に新しいことでしょう。特定の領域においては、AI が人間を超え始めている一例ともいえます。

昨今のインターネットやスマートフォンをはじめとするモバイルデバイスの急速な普及に伴い、データの爆発的な増加と情報処理技術の革新が起き、これまで不可能とされてきた規模のデータ活用が可能となりました。さらに 2020 年 4 月にサービスを開始した 5G 回線の登場によってモバイルデータ通信の高速・大容量化が加速し、収集できるデータも拡大しました。

技術進歩だけでなく、データ活用はすでに私たちの生活に身近な存在となっています。ICT 産業に限らず、自動車、健康医療、インフラ、エンターテインメント、農業……あらゆる産業において実用化が始まっています。こうした膨大なデータのリアルタイム処理も可能になることから、データ活用の重要性がより一層高まっていくことは明らかです。

私たちは今、「データを活用して新しい価値を生み出す」力が必要とされる世界に生きています。これまでは既存の市場やルールのなかで切磋琢磨し活動することがモノ・カネを生み出すことにつながってきましたが、これからは、枠組みを超えた新しい価値を創出・創造することが富へとつながる時代を迎えます。

その新しい価値を生む鍵となるのがデータや AI です。私たちはデータをどのように扱い、どのような価値を、どのような社会を生み出していくのかをデザインし、実現させるスキルを磨かなければなりません。

ひとりでも多くの人々がデータを味方につけ、新たな社会を描いてほしい。データサイエンティストスキルチェックリスト/タスクリストはそんな思いを込めて、データ利活用のために必要とされるスキルを体系化しています。学生、社会人、データ活用の課題に直面している人、そして今は縁がないと思っている人にも、本書で取り上げているスキルが新時代を生き抜く一助になれば幸いです。

# 目次 CONTENTS

はじめに Preface .....	1
--------------------	---

## Part 01

<b>WHY データサイエンス?</b> .....	5
社会背景 .....	6
身近にあるデータサイエンス .....	8

## Part 02

<b>データサイエンティストに必要なスキルセットとタスク</b> .....	11
スキルチェックリストとは? .....	12
タスクリスト .....	21
スキルチェックリスト / タスクリストの構造 .....	22

## Part 03

<b>スキルチェックリストの読み方</b> .....	23
全体像を理解する .....	24
スキルチェックリストの見方 .....	30

## Part 04

<b>タスクリストの読み方</b> .....	33
全体の構造を理解する .....	34
タスクリストのカラム説明 .....	36
タスクリストにも流れがある .....	38



Part 05

<b>スキルチェックリスト・タスクリストの使い方</b> .....	43
データサイエンティスト初心者・目指している人 .....	44
データサイエンティスト実務家 .....	45
データサイエンティストチームリーダー プロジェクトマネージャー ...	46
外注計画担当者 .....	47
人事採用担当者・エージェント .....	48
学校の先生 .....	49
学生（学ぶ側の人） .....	50
育成制度を作ろう .....	51

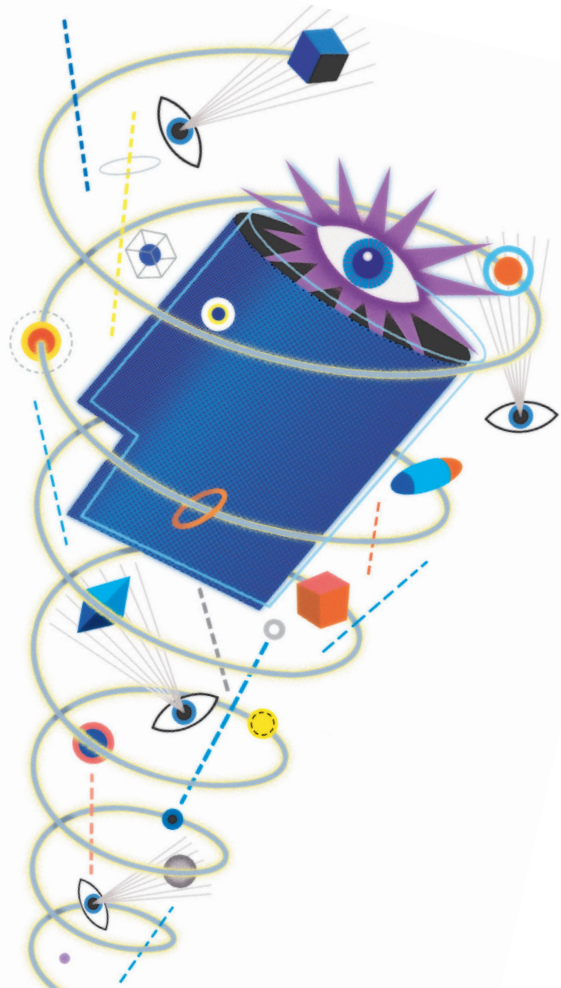
Part 06

<b>リストにない定義や考え方</b> .....	55
ドメインスキル .....	56
文章に込められた意味 .....	57
進化の速い分野 .....	60
ビジネスや事業への関わり方 .....	64
お役立ち情報 Appendix .....	66



Part 01

# WHY データサイエンス？



## ● 社会背景

### 社会的ニーズの高まりから生まれた

近年の技術革新によって、私たちが手にできるデータは爆発的に増大しました。そうしたデータを応用したいという世の中のニーズの高まりに呼応して活用できる人材が広く求められるようになり、登場したのが「データサイエンティスト」という職業です。**データサイエンスとは、アルゴリズムや統計などといった情報科学系の理論を活用してデータを分析し、有益な知見を見出すことを追究する新しいアプローチ**です。

この10年ほどの間にデータサイエンスによるデータ活用の手法や成果は飛躍的に発展し、多くのデータサイエンティストが活躍するようになりました。多くの企業や分野でデータサイエンティストが求められ、人材不足が懸念される状況にもなっています。その一方で、「データサイエンティストの仕事も将来的にはAIに取って代わられるのでは……」という声も聞かれるようになりました。もちろん、人間よりもAIが計算・処理した方が速度も速く精度も高い識別や予測などの領域は、AIに代替されていくものでしょう。しかし、**どのようにデータを使って社会にどのような価値を生み出すかをデザインし実行していくことは、人間にしかできない仕事**です。

AIの技術・研究のために、世界各国から優秀な人材が研究機関や企業に集められ、莫大な投資が行われています。日本でも、2019年6月に「AI戦略2019～人・産業・地域・政府全てにAI～」が発表され、デジタル社会の基礎知識（いわゆる「読み・書き・そろばん」的な素養）として、すべての国民が「数理・データサイエンス・AI」を学ぶことが目標として掲げられました。このAI戦略2019を受け、高校では2022年度から「情報I」が必須科目になるとともに、大学・高専では、文系理系を問わず全学部向けの教育として「数理・データサイエンス・AI教育」の展開が始まっています。

データや AI を活用することで何ができるのか、どうすればより社会に役立てられるのかを考え、**新しいデジタル社会を実現していくためには、データサイエンスに関するスキルがとても重要**になります。同時に、スキルレベルを測る手段の重要性も徐々に増してきました。

データサイエンティスト協会では、そうした社会的ニーズに応えるために、「データサイエンティスト検定 リテラシーレベル」という検定試験を実施しています。

## ●身近にあるデータサイエンス

### すべての人に関係するデータサイエンス

---

データサイエンスと聞くと、「難しい」「理系にしかわからないもの」と感じる人も多いかもしれません。しかし**データサイエンスは、すでに私たちの生活の中でとても身近な存在**となっています。

インターネット通販で商品を選んでいるとき、「あなたへのおすすめ商品」が表示されることがありませんか。これは、過去に同じ通販サイトに訪れた人たちの行動履歴や商品データから、あなたと似たパターンのショッピングをしている人や、似た商品を探し出すことで、あなたが気に入ってくれそうな商品を探して表示しています。

また、近年よく耳にする**自動運転も、データサイエンスの代表的な実用事例**の一つです。自動運転は、走行中の周辺の様子をカメラやセンサーで捉え、そのデータを分析することで周辺の危険を予測し、ハンドルやブレーキを操作してくれます。

最近ではスポーツの領域でも、データサイエンスの活用事例に触れることが多くなりました。メジャーリーグの中継では、打席に立つバッターが映し出されている場面で、ストライクゾーンが画面上に枠となって表示されています。これもリアルタイムで映像を解析することで、どこがストライクゾーンになるのかを判定しています。2021年の東京オリンピック・パラリンピックにおいてもリアルタイムの判定システムや選手の動きの立体的な映像解析、ボールの弾道解析などは記憶に新しいところです。

また、コロナ禍において、日々の感染者数・死者数・重症者数が報道されていますが、これらのデータの多くはオープンデータとして公開、利用されています。こうしたデータとGPSなどのデータを組み合わせて、外出自粛率の算出や、接触頻度の見える化などを行うことも、データサイエンスの活用事例の一つです。

一方で「フェイクニュース」などの言葉を最近では聞かれたことがある

と思います。機械学習の技術を利用し、SNS 上で実在しない人物や偽の記事を氾濫させ、社会的な混乱やプライバシーの侵害などの問題を招いています。データサイエンティストとしてデータや技術をどう扱うのか？どのように社会に還元するのか？が真摯に問われている一面といえます。

このように、**データサイエンスのスキルが活用される業界やシーンは非常に幅広く、その裏では多くのデータサイエンティストが活躍しています。**また活用シーンが拡大することで、今は縁のない業界や職種の方でも、当たり前のようにデータを活用する世界が目の前までできています。

**データサイエンスのスキルは、あらゆる業種や職種の方にとっても重要な基礎素養**となっていくことでしょう。

## データサイエンティストとは？

データサイエンティストとは、「データサイエンス力、データエンジニアリング力をベースにデータから価値を創出し、ビジネス課題に答えを出すプロフェッショナル」です。

ここでの「ビジネス」とは企業の営利活動だけではなく、社会の役に立つ意味のある活動全般を指します。「プロフェッショナル」とは、体系的にトレーニングされた専門的なスキルをベースに顧客（お客様、クライアント）に価値を提供し、その対価として報酬を得る人です。

**図1** データサイエンティストとは、裏打ちされたスキルを駆使してデータを分析し、ビジネス課題の回答を導き出すプロフェッショナルを指します

### スペシャリスト

- 何らかの分野に特化した人

### エキスパート

- 何らかの分野について、体系的で秀でた知識とスキルを持っている人

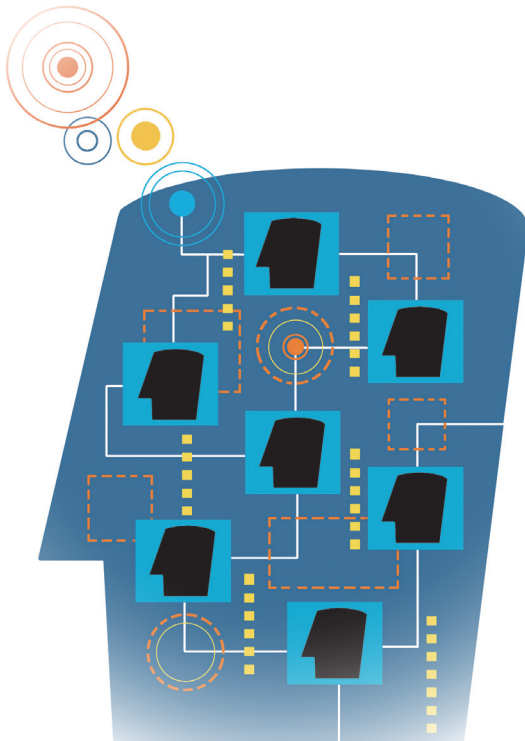
### プロフェッショナル

- 体系的にトレーニングされた専門性の高いスキルを持ち、
- それをベースに顧客（お客様、クライアント）にコミットした価値を提供し、
- 結果に対して認識された価値の対価として報酬を得る人



Part 02

# データサイエンティストに必要なスキルセットとタスク



## ● スキルチェックリストとは？

### 必要なスキルを分野別、体系的に一望・俯瞰

---

Part01 で紹介したとおり、データサイエンスを駆使するデータサイエンティストは、これからの時代にとって欠かせない存在となり、データサイエンティストを志す人やプロフェッショナルとしてのデータサイエンティストを育成する企業が増えています。

しかし、データサイエンティストになりたい！ 当社でもデータサイエンティストを育成したい！ と考えたとき、まず何から始めればよいのでしょうか。

データサイエンティストのスキル（＝できること）を一望、俯瞰できれば、データサイエンティストを目指す人の一助になるに違いないという意識のもとで作成したのが「スキルチェックリスト」です。

スキルチェックリストには、データサイエンティストとして活躍するための必要なスキル（スキル項目）を、分野ごとに体系的にまとめています。スキルチェックリストを見れば、データサイエンティストを目指す人やデータサイエンティストを育成したい企業にとって、勉強や教育の目標とすべきスキルが一目でわかります。

それぞれのスキル項目には、「スキルレベル」を定義されていますので、基礎から高難易度まで、必要とされるスキルレベルが確認できます。

## スキルチェックリストのねらい

---

「データサイエンティストにできること(=スキル)」をスキルチェックリストとしてまとめることに至った背景には、一般社団法人データサイエンティスト協会・スキル定義委員会が共有してきた以下のような問題意識がありました。

- ・「データサイエンティスト」という言葉だけがひとり歩きし、本質が十分に理解されていない(または誤解されている)。
- ・「データサイエンティスト」に求められるスキルの幅の広さやレベルの高さが十分に理解されていない。
- ・「データサイエンティスト」を育成するための体系的な指標がない。

これらを踏まえて、「データサイエンティスト」の仕事やスキルについてもっとくわしく、より多くの方々に知ってもらうためにまとめたのがスキルチェックリストです。

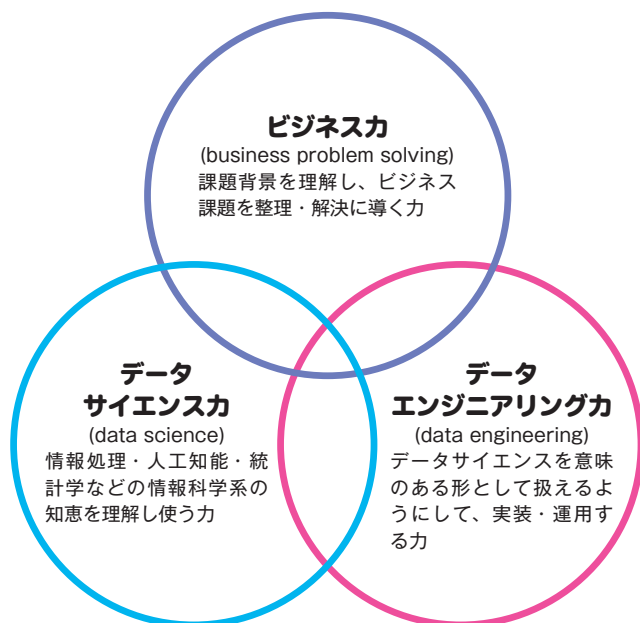
データサイエンティストのできることを単に紹介するのではなく、企業や組織のなかで活躍し、これからの社会、これからの日本を変えていけるような真のデータサイエンティストを数多く生み出したいという願いも込められています。

## 3つのスキルセット

データサイエンティストに求められているスキルセットは、以下の3つです。

- ・ **ビジネスカ (business problem solving) :**  
課題背景を理解し、ビジネス課題を整理・解決に導く力
- ・ **データサイエンスカ (data science) :**  
情報処理・人工知能・統計学などの情報科学系の知恵を理解し使う力
- ・ **データエンジニアリングカ (data engineering) :**  
データサイエンスを意味のある形として扱えるようにして、実装・運用する力

図2 データサイエンティストに求められるスキルセット



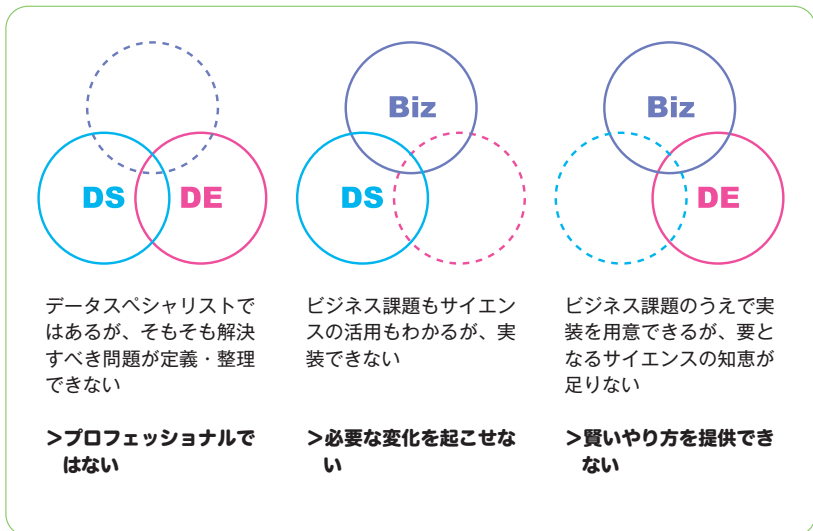
### 3つのスキルセットはいずれも欠けてはいけない

3つのスキルセット（ビジネスカ、データサイエン্সカ、データエンジニアリングカ）はとても重要なファクターで、どれか一つが欠けてもデータサイエンティストとして十分な力を発揮できません。

ビジネスカに欠ければ、そもそも解決すべき問題が定義、整理することができません。データエンジニアリングカが乏しければ、ビジネス課題についてのサイエン্সの活用はわかっても、肝心の実装（実際に利用できるシステムを仕立て上げること）ができません。データサイエン্সカがなければ、ビジネス課題の上で実装を用意できても、要となるサイエン্সの知恵が足りないため、データの力を活かすことができません。

スキルセットは、いわば三位一体なのです（**図3**）。

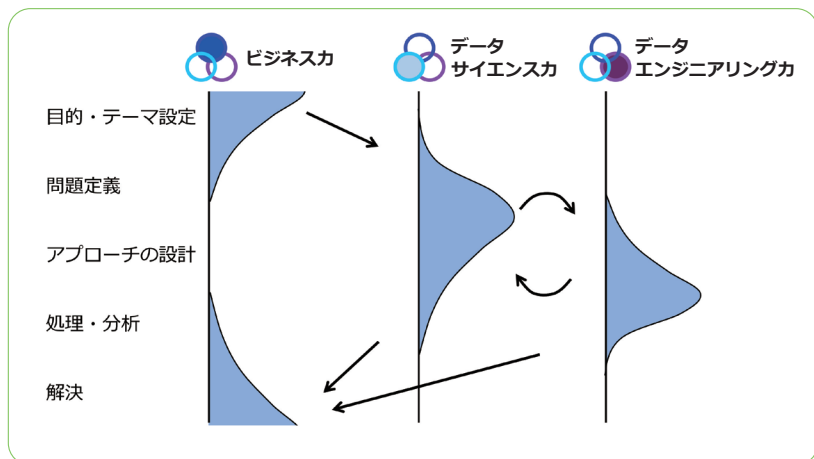
**図3** いずれのスキルセットが欠けてもデータスペシャリストとしての力は発揮できない



3つのスキルセットはデータサイエンティストにとって欠かせないファクターですが、それぞれの重要性は課題解決のフェーズによって変わってきます（図4）。

目的・テーマ設定や最後の解決などのフェーズでは、関係者を説得・調整したり、説明したりするようなビジネス力が強く求められます。定義された課題に対してどのようなアプローチで立ち向かうのかを計画するには確かなデータサイエン্স力が必要です。そして、実際の分析を進めるためには、データエンジニアリング力が要求されます。

図4 問題解決のフェーズによって要求されるスキルセットは変わる



### よくある誤解 一人でできたらスーパーマン!?



データサイエンティストに要求されるスキル定義は、必ずしも一人ですべてのスキルをこなすことを想定したものではありません。スキルセットごとに得手不得手もあるでしょう。最初からすべてを淀みなくこなせたらそれこそスーパーマン!?! かもしれませんね。

得意なスキルを伸ばすことで問題解決に貢献できます。全体を俯瞰してコーディネートする人材は必要ですが、個々のスキルに秀でたメンバーでチームを組み、実務にあたることも現実には多くなるでしょう。こうして経験を積むことは、一人前のデータサイエンティストになるための糧になるのです。

## スキルレベルの4段階

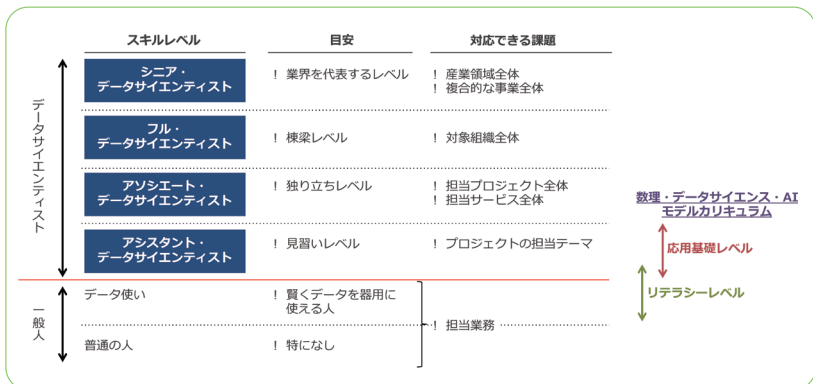
データサイエンティストにとって必須であるスキルセットに対して、それぞれに4段階のスキルレベルがあります。

1. 業界を代表するレベル : Senior Data Scientist
2. 棟梁レベル : Full Data Scientist
3. 独り立ちレベル : Associate Data Scientist
4. 見習いレベル : Assistant Data Scientist

スキルチェックリストは、見習いレベル (Assistant Data Scientist) ~ 棟梁レベル (Full Data Scientist) までの判定に用います。

また、近年、高校や大学でデータサイエンス授業の必修化の動きがありますが、数理・データサイエンス教育強化拠点コンソーシアムが発表した「数理・データサイエンス・AI モデルカリキュラム」(以下、モデルカリキュラム) で示されたリテラシーレベル、応用基礎レベルとスキルレベルの対比については、**図5**の右に示すように、モデルカリキュラムのリテラシーレベル、応用基礎レベルはそれぞれ見習いレベルとその下のレベルに相当するスキルであると位置づけています。

**図5** データサイエンティストのスキルレベル



## モデルカリキュラムとは？

AI 戦略 2019 に基づき、数理・データサイエンス教育強化拠点コンソーシアムでは、数理・データサイエンス・AI 教育の全国展開を目指し、リテラシーレベル（50 万人 / 年目標）および応用基礎レベル（25 万人 / 年目標）のモデルカリキュラムを公開しています。

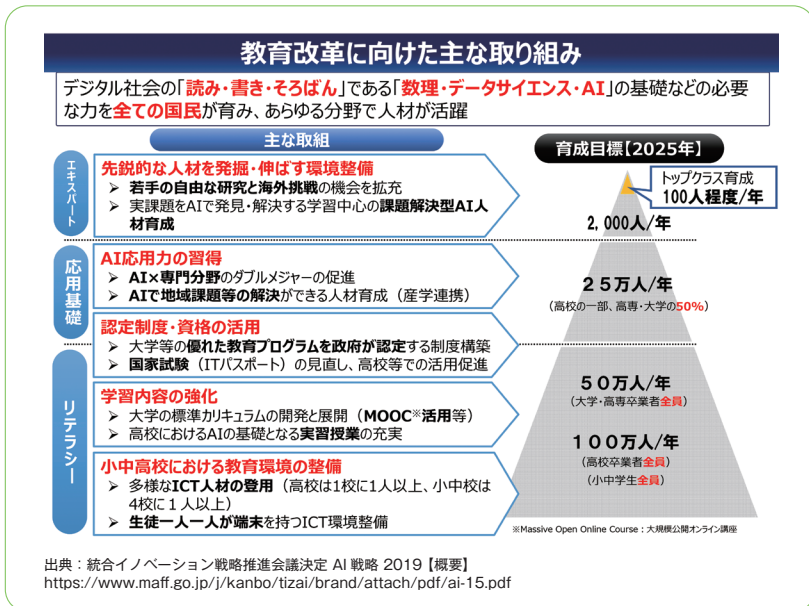
### 数理・データサイエンス・AI（リテラシーレベル）モデルカリキュラム ～データ思考の涵養～

[http://www.mi.u-tokyo.ac.jp/consortium/pdf/model\\_literacy.pdf](http://www.mi.u-tokyo.ac.jp/consortium/pdf/model_literacy.pdf)

### 数理・データサイエンス・AI（応用基礎レベル）モデルカリキュラム ～AI×データ活用の実践～

[http://www.mi.u-tokyo.ac.jp/consortium/pdf/model\\_ouyoukiso.pdf](http://www.mi.u-tokyo.ac.jp/consortium/pdf/model_ouyoukiso.pdf)

図 6 AI 戦略 2019（抜粋）





## スキルチェックリストの変遷 (Ver.1 ~ Ver.3)

データサイエンティストに求められるスキルは、技術や産業の進化に伴い変化していきます。新たな技術が生まれたり、重要な技術が新技術に置き換えられて必要性がなくなったり、かつては高難度だったものが低難度に変化したりすることもあります。

特にデータサイエンス領域は進化・深化が早いいため、これまでは全スキルチェック項目の更新を2年に1回実施して最新のスキルを反映してきました。

2017年 (Ver.2) に見直したおもな内容は以下の通りです。

- 機械学習、深層学習時代に対応したスキル項目の見直し・追加
- 「画像・動画処理」「音声/音楽処理」のカテゴリ・スキル項目の見直し
- 「知財」スキルカテゴリの追加

2019年 (Ver.3) に見直したおもな内容は以下の通りです。

### 【データサイエンスカ】

- 「統計数理基礎」を「基礎数学」とスキルカテゴリ名変更し、線形代数や微積分に関するスキル項目を追加
- 「予測」関連のスキルカテゴリを「回帰/分類」「前処理」「モデル設計」「評価」へと再分類し、精緻化

### 【データエンジニアリングカ】

- クラウド環境を用いた分析構築スキル項目の追加
- セキュリティ関連スキル項目の強化  
(例：個人情報や GDPR 等を考慮したデータ管理スキル)
- 近年の技術動向反映  
(例：p 値の限界、特徴量エンジニアリング DL ライブラリ活用)

### 【ビジネスカ】

- 分析プロジェクトの初期段階で実務に関するスキル項目を強化  
(例：「着想・デザイン」「契約権利保護」等)
- 事業実装に関連するスキルカテゴリ「事業への実装」「分析評価」の見直し・追加

## スキルチェックリストの変遷 (Ver.4)

2021年 (Ver.4) に見直したおもな内容は以下の通りです。

### 【全体】

- モデルカリキュラムとの整合性の調整
- AI活用や機械学習のシステム運用、アルゴリズム、それに伴う倫理課題に関するスキル追加
- ライブラリ活用やクラウドサービスの一般化に伴うスキルの更新

### 【データサイエンスカ】

- 基礎数学の重要性見直し
- 学習と予測カテゴリの明確化、潮流への対応
- 近年、重要度の増したスキル項目追加

### 【データエンジニアリングカ】

- 新規カテゴリ「AIシステム運用」を追加
- 環境構築技術等、技術トレンドへの対応
- リテラシーレベルのプログラミングスキル追加
- HadoopやScala、GPUなど一般化技術の一部削除

### 【ビジネスカ】

- AI活用時のビジネス視点を強化
- プロジェクトの多様なフェーズに対応するスキル
- DSから「分析アプローチ」スキル移設
- マネジメントスキルをプロジェクトと組織に2分割

## ●タスクリスト

「データサイエンティストにできること＝スキル」をまとめたスキルチェックリストに対して、「データサイエンティストが行うこと＝タスク」をまとめたものが「タスクリスト」です。

タスクリストは、スキルチェックリストを補完するもので、実際にどのようなプロセス（業務）においてスキルチェックリストに記載されているスキルが必要とされるかを整理したものです。

データサイエンスを駆使して価値を生み出す活動の流れそのものがよくわかっていない、自分はデータサイエンスのプロジェクトのなかでどのような役割を果たしたいのかがはっきりしない……といったときに役立つのがタスクリストです。そのような状況になったら、まずタスクリストを見て業務のなかで自身が担当するタスクは何かを確認します。

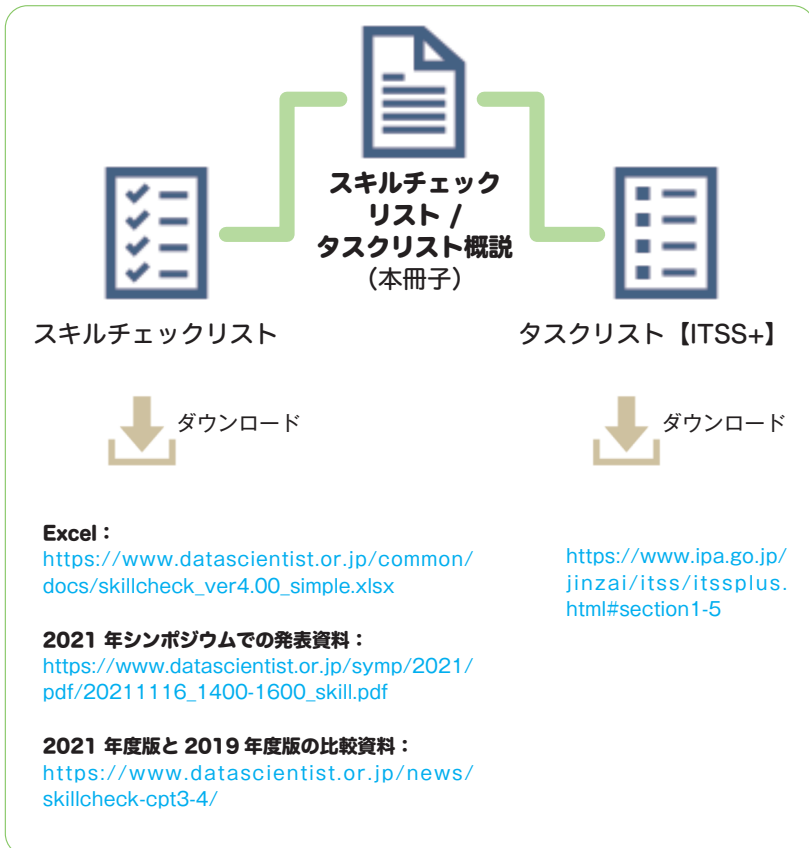
その後でスキルチェックリストを参照して、担当するタスクを行うには、どのスキル項目を身に付ける必要があるのかを確認するといった使い方を想定しています。

## ● スキルチェックリスト / タスクリストの構造

本書も含めたドキュメントの構造を下図に示しました（図 7）。

独立行政法人情報処理推進機構（IPA）は、スキル標準の新領域「ITSS+（データサイエンス領域）」整備を一般社団法人データサイエンティスト協会スキル定義委員会とともに取り組んでいます。スキルチェックリスト、タスクリストは、相互のウェブサイトで公開しています。

図 7 関連するドキュメントの構造



Part 03

# スキルチェックリストの 読み方



## ●全体像を理解する

### スキルチェックリストの構成

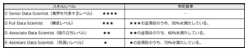
スキルチェックリストは、以下のパートで構成されています（表 1）。

- ・チェックリストの説明
- ・スキルレベル定義
- ・スキル一覧（各スキルセット）
- ・参考資料

まず「チェックリストの説明」シートと「スキルレベル定義」シートでスキルレベルの判定基準やスキルカテゴリの一覧、スキルレベルの人物像などを確認します。続いて、各分野のスキル一覧を眺めることで、全体感が把握しやすくなります。

なお、スキルチェックリストは適宜（今までは2年ごと）更新されるため、手元にあるものが最新版かどうか必ず確認しましょう。

表 1 スキルチェックリストの全体構成

シート名	記載内容
チェックリストの説明	<ul style="list-style-type: none"><li>・スキルレベルの判定基準</li><li>・スキルカテゴリの一覧</li></ul> 
スキルレベル定義	<ul style="list-style-type: none"><li>・分野ごとのスキルレベルの人物像</li></ul> 
ビジネスカ データサイエンスカ データエンジニアリングカ	<ul style="list-style-type: none"><li>・スキルセットごとのスキル一覧</li></ul> 
参考資料 (データ可視化)	データサイエンスカの「データ可視化」に関する参考資料 (2015年11月20日時点)  右図は株式会社 ALBERT WEB サイト <a href="https://blog.albert2005.co.jp/2014/12/11/">https://blog.albert2005.co.jp/2014/12/11/</a> 

## スキルレベルの判定基準

「チェックリストの説明」シートには、スキルレベルの判定基準が記載されています。各スキルレベル達成のためには、必須スキルをすべて満たしたうえで判定基準を達成する必要があります（**図 8**）。

**図 8** スキルレベルの判定基準

「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）～ 棟梁レベル（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 (棟梁レベル) ★★★★	★★★★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。

**図 8** の「独り立ちレベル」になるための条件を例にあげると、**表 2** で示した条件をすべて満たす必要があります。判定によっては、データサイエンス力は「独り立ちレベル」で、ビジネス力は「見習いレベル」ということもあります。

**表 2** 独り立ちレベルに達するための条件（データサイエンス力の例）

条件 1	データサイエンス力の★の全項目のうち、70% を満たしている
条件 2	データサイエンス力の★★の全項目のうち、60% を満たしている
条件 3	データサイエンス力の★の「必須スキル」をすべて満たしている
条件 4	データサイエンス力の★★の「必須スキル」をすべて満たしている







図 11 データサイエンスカルのスキルカテゴリ

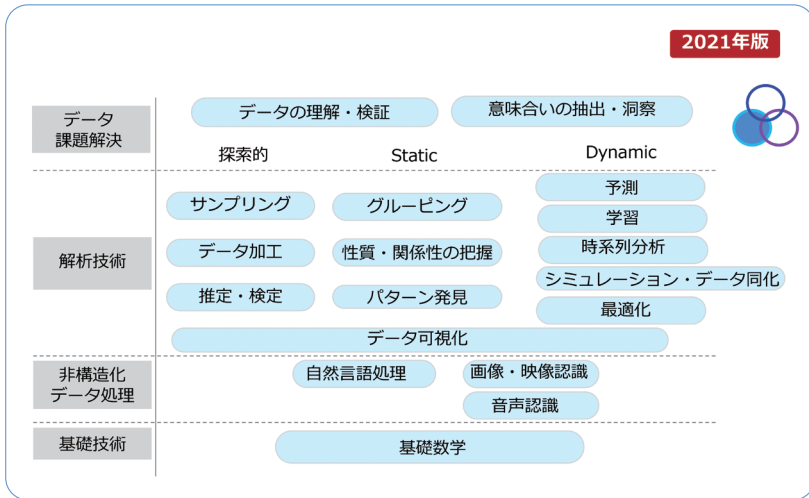
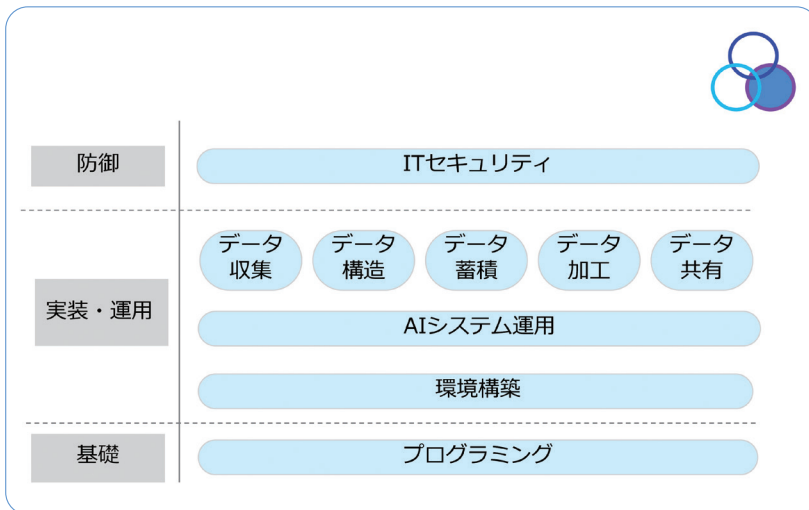
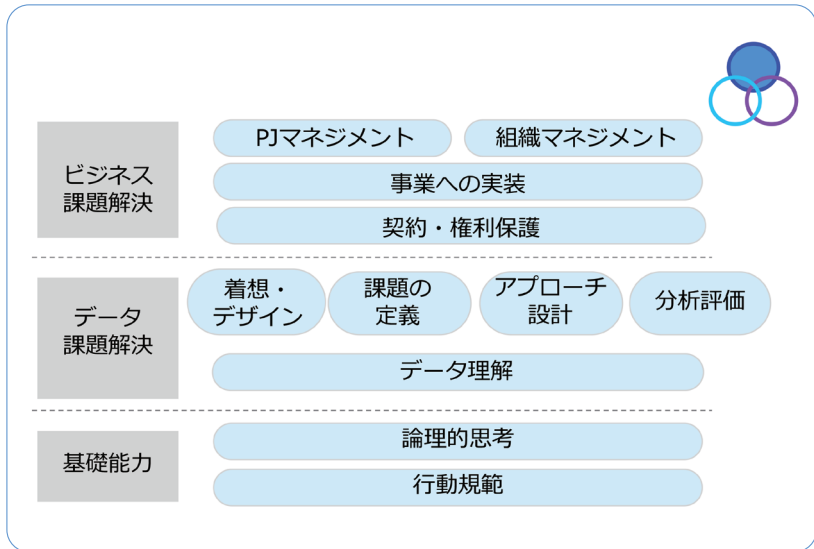


図 12 データエンジニアリングカルのスキルカテゴリ



## 図13 ビジネスカのスキルカテゴリ



# ● スキルチェックリストの見方

## スキルチェックリストの各カラムについて

スキルチェックリストは複数の項目で分類されています (図 14)。まずは、「スキルカテゴリ」、「サブカテゴリ」をざっと眺めて、気になる項目をチェックしてみましょう。

図 14 データサイエンスカのスキルリストからの抜粋

スキルの中分類 (詳細は Part03「全体像の理解」P.27 へ)		スキルのレベル (詳細は Part03「全体像の理解」P.25 へ)		他分野にまたがるスキルには * を付与		必須スキルには ○ を付与		
NO	SubNo	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
125	4	データ加工	*	データ加工	名義尺度の変数をガミ-変数に変換できる			
126	5	データ加工	**	データ加工	加工データに不具合がないか確認方法を設計し、検証できる (合計値の照合や、分布の確認など)	*		○
127	6	データ加工	***	データ加工	加工データの統計的な傾向によって不具合の早期発見ができること、統計的観点で次ステップの解析に繋がるデータであるかを評価できる			○
128	7	データ加工	*	特徴量エンジニアリング	数値データの特徴量化 (二値化/離散化、対数変換、スケーリング/正規化、交互作用特徴量の作成など) を行うことができる	*		
129	8	データ加工	**	特徴量エンジニアリング	カテゴリデータの特徴量化ができる (2値-変数化、特徴量ハッシュ[カテゴリ取りが多い場合の節約手段]、ヒカウワンエンコーディング、バグクワフ、最小ワウツトクック[レアなカテゴリのみを処理]など)	*		
130	9	データ加工	**	特徴量エンジニアリング	k-means、k近傍法、t-SNEなどを用いた、類似度または非類似度に着目した特徴量を作成できる	*		
131	10	データ加工	**	特徴量エンジニアリング	局所的な意味合いをもつ特徴量を適切な周期の三角関数を用いて変換できる	*		
132	11	データ加工	**	特徴量エンジニアリング	時系列データに対し周波数やトレンドなどを考慮して、階差、移動平均、移動標準偏差、移動総大、移動最小、シフトなどの特徴量を作成できる	*		
133	12	データ加工	**	特徴量エンジニアリング	複数のデータセットから、結合や集計などを行い合わせ、有効な特徴量を作成できる (ユーザーごとの過去1ヵ月間の購買金額合計など)	*		○
134	13	データ加工	***	特徴量エンジニアリング	ドメイン知識に基づき洞察から有効な特徴量を効率的に作成できる (類似商品との価格差、借入額と返済額の比率など)	*		
135	14	データ加工	***	特徴量エンジニアリング	有用と認められるデータを新たに調査し、既知のデータと組み合わせることで有効な特徴量を作成できる (気象特異なカレンダー、地理空間的な類似度を考慮した集計値など)	*		
136	1	データ可視化	*	方向性定義	データの性質を理解するために、データを可視化し、極めて考えることの重要性を理解できる	*	○	
137	2	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の場合から、ビッグデータ中の重要な関連性をダイナミックに表示する場合など)	*		
138	3	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、方向性を示すことができる	*		
139	4	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、プロットが想像に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 (ビッグデータ中の重要な関連性をダイナミックに表示する、疑問に入りきりに問いに対して答えを出すなど)	*		
140	5	データ可視化	*	輸出し	配布意図などの輸出しにおいて、目的やデータに応じて縦軸・横軸の候補を適切に洗い出せる	*		○
141	6	データ可視化	*	輸出し	構み上げ縦軸グラフでの属性の選択など、目的やデータに応じて適切な層化 (比較軸) の候補を提出できる	*		○
142	7	データ可視化	**	輸出し	抽出したい意味におよびたい軸・層化の程度、階層を考慮し軸のきざみや層化方法を選択できる	*		
143	8	データ可視化	**	輸出し	膨大な属性を持つテーブルが目的に有用な属性を選択できる	*		
144	9	データ可視化	**	輸出し	非構造データが分析の軸にならざる様子を指摘し、付加すべき属性候補を適切に出せる	*		
145	10	データ可視化	*	データ加工	サンプリングやサンブル平均によって過剰なデータ量を減らすことができる	*		
146	11	データ可視化	*	データ加工	読み取りにくい特徴を効果的に可視化するために、統計量を使ってデータを加工できる	*		
147	12	データ可視化	**	データ加工	データの持つ特徴量の観点で、高次元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	*		○
148	13	データ可視化	***	データ加工	高次元の非線形な (高次の曲線、渦状の分布などの) データであっても、適切に1〜3次元のデータに変換して、特徴 (データの総分数量および各データの位置関係) を構わずに可視化できる	*		
149	14	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、ネットワーク処理をうまく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などして適切に可視化できる	*		
150	15	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) なるため、処理しきれず描画できない規模のデータに対しては、適度なデータや情報の抽出 (即引き)、クワクワ分析などにより可視化した状態にデータを加工できる	*		

表3 スキルチェックリストの掲載項目

No	スキルセット全体の通番です。
Sub No	スキルカテゴリ内での通番です。
スキルカテゴリ	スキルを目的や技術のカテゴリに応じて分類したものです。
スキルレベル	各スキルが該当するレベルを見習いレベル(★)、独り立ちレベル(★★)、棟梁レベル(★★★)で表しています。
サブカテゴリ	スキルカテゴリを詳細に分類したものです。
チェック項目	スキルに求められる内容が記載されています。
DS、BZ、DE	他のスキルセットと分野がまたがる傾向にあるスキルの場合、*印がつきます。たとえばビジネス力中心のキャリアを目指す場合には、ビジネス力のスキルだけではなく、他分野（データサイエンス力、データエンジニアリング力）でビジネス力に*がついているスキルも確認してみましょう。 DS・・・データサイエンス力 BZ・・・ビジネス力 DE・・・データエンジニアリング力
必須スキル	スキルレベル達成のために必須のスキルです。共通に必要なスキルにのみ○がつけられています。必須スキルを一つでも満たさない場合は、当該スキルレベルを達成できません。

### 必須スキル以外にも重要なスキルはある！



たとえば、SQL（データベースの操作で使用する言語）はデータエンジニアリングの基本であり極めて重要なスキルではありますが、分析内容（自然言語処理、画像認識など）によっては使用されないこともあります。また、近年ではクラウドコンピューティングに関する知識や技術も重要なスキルですが、自社の情報セキュリティルールなどにより使用が禁じられている場合もあります。

そうした背景から、SQL やクラウド関連技術など、重要ではあるものの一部のスキルは必須スキルと定めていないものがあります。必須スキルではなくても、近年の技術動向を踏まえ、重要だと思われるスキルは身に付けるようにしましょう。



## スキルチェックリストの読み方

スキルチェックリストは、さまざまな読み方ができます。全体の俯瞰やスキルアップに向けた流れを意識した読み方、未知の領域を学ぶための知らない用語の拾い読み……、用途に合わせて読んでみましょう。

### ●カテゴリ単位、サブカテゴリ単位の俯瞰

カテゴリ単位、サブカテゴリ単位で全体を俯瞰し、自分に関係するスキルを絞り込むこともできますでしょう。

### ●同一カテゴリのスキルレベルの流れ

同一カテゴリにおいて、「★」→「★★」→「★★★」といったスキルレベルのアップを目指した読み方もできますでしょう。

### ●同一カテゴリ、同一サブカテゴリ、同一スキルレベルの流れ

同一カテゴリ、同一サブカテゴリ、同一スキルレベルで、どのようなスキルがあるか、スキルの幅を広げる読み方もできますでしょう。

### ●未知の用語のチェック

スキルチェックリストは、見習いレベル～棟梁レベルまで必要とされるスキルに関する用語の宝石箱です。関連するカテゴリ、サブカテゴリ内で知らない用語を見つけたら、スキルを獲得するチャンスと考えましょう。

Part 04

# タスクリストの読み方

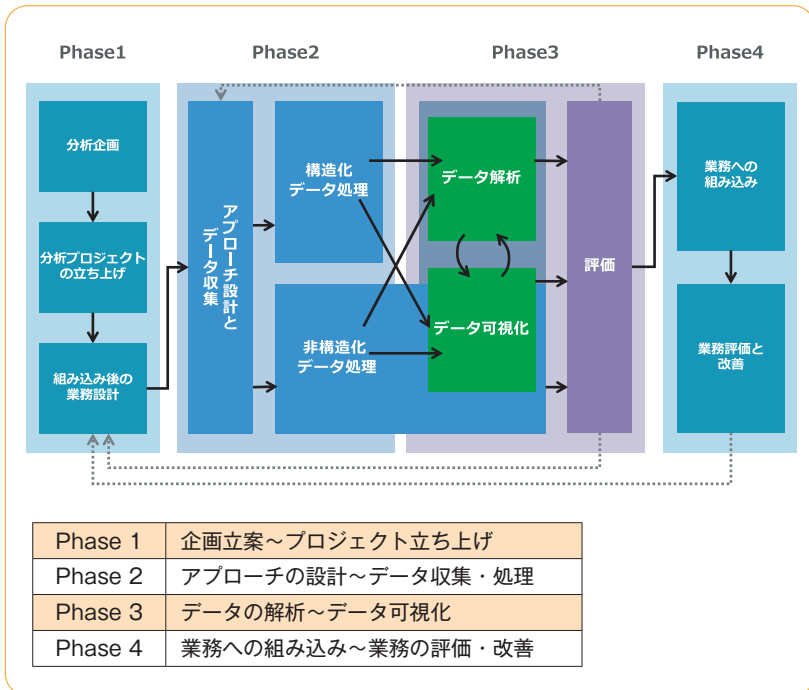


## ●全体の構造を理解する

### データサイエンティストの仕事・業務

データサイエンティストが実際に行う業務は、大別すると4つのフェーズで構成されます（**図15**）。Phase 1では、企画立案からプロジェクトの立ち上げを行います。Phase 2は、プロジェクトについてのアプローチの設計から、それにまつわるデータを収集し、処理するフェーズとなります。Phase 3では、前フェーズで下ごしらえしたデータの解析および可視化を行います。そして最終段階となるPhase 4では、業務への組み込み（実装）とともに、業務自体の評価・改善などを実施します。

**図15** タスクリストにおける4つのフェーズ



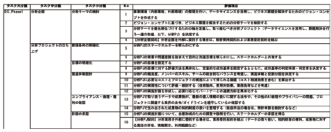
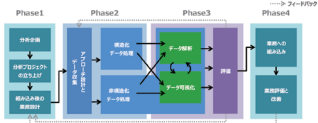



## 業務タスクはタスクリストとタスク構造図で確認

データサイエンティスト業務のタスクを確認する際には、「タスクリスト」と「タスク構造図」の両方を用います。

タスクリストには、データサイエンティストが担う一連のタスクが記載されています。タスク構造図には、データサイエンティスト業務の主要な流れと改善のために行うタスクの戻りについてが記載されています。

図 16 タスクリストとタスク構造図

資料	内容	イメージ
タスクリスト	データサイエンティストが担う一連のタスクを記載	
タスク構造図 (中分類)	データサイエンティスト業務の主要な流れ (タスク中分類レベル) とタスクの戻りについて記載	
タスク構造図 (小分類)	データサイエンティスト業務の主要な流れ (タスク小分類レベル) とタスクの戻りについて記載	

# ●タスクリストのカラム説明

## タスクリストを確認しよう

タスクリストは、カテゴリー別に分類し、140 項目のタスクで構成されます(表4、図17)。まず、「タスク中分類」と「タスク小分類」を眺めながら、データサイエンティスト業務にどのようなタスクがあるか確認してみましょう。

興味を持ったタスク分類があれば、「評価項目」欄でどのようなタスクを行うのかということを確認できます。

表4 スキルチェックテストの各項目

タスク大分類	データサイエンティスト業務におけるフェーズを示す
タスク中分類	データサイエンティスト業務におけるタスクの分類を示す
タスク小分類	タスク中分類の中のさらに詳細な分類を示す
No	タスクリスト上の「評価項目」(下行参照)の通し番号
評価項目	各々のタスクを記述した項目

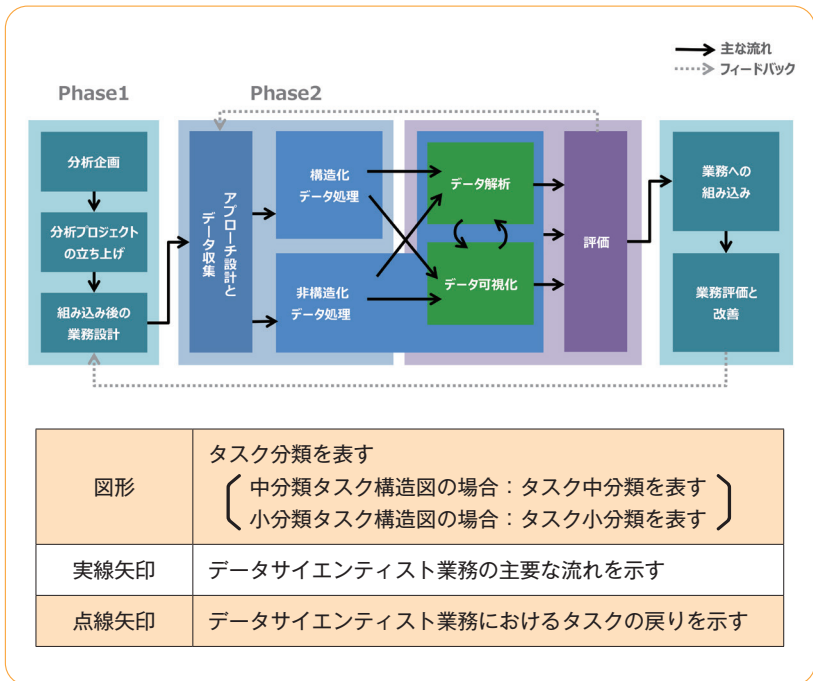
図17 タスクリストの抜粋

タスク大分類	タスク中分類	タスク小分類	No	評価項目
DS_Phase1	分析企画	分析テーマの検討	1	事業環境(内部環境、外部環境)の整理を行い、データサイエンスを活用し、ビジネス課題を解決するためのビジョン・コンセプトを作成する
			2	ビジョン・コンセプトに基づき、ビジネス課題を解決するための分析テーマを検討する
			3	分析テーマを単体で立つための軸を定義し、取り届くべき分析プロジェクト(データサイエンスを活用し、課題解決を行う一連の組織、以下、分析P)を決定する
			4	[分析企画案] 分析企画を外部に委託する場合は、秘密保持規約および業務委託契約を結ぶ
	分析プロジェクトの立ち上げ	前提条件の明確化	5	分析Pのステークホルダーを明らかにする
			6	分析Pの背景や問題意識を踏まえて目的と到達目標を明らかにし、ステークホルダーと共有する
		目標の明確化	7	分析Pの目標を設定する
			8	分析Pの目標に対する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、成功基準の判定時期・判定者を決定する
		推進体制設計	9	分析Pの難易度、メンバーのスキル、チームの総合的なバランスを評価し、実働体制と役割分担を決定する
			10	分析Pに必要なコストとプロジェクトの実現によって得られる価値(コスト削減効果を含む)を算出する
			11	分析Pの実現性について評価・検討する(技術的、費用対効果、業務負荷などを考慮)
			12	分析Pの実施計画を作成し、必要に応じてパートナーとの連携方針を具体化する
		コンプライアンス、倫理、権利の確保	13	分析Pで取り扱うデータや成果物が、顧客の個人情報保護に関する法令や、その他のAI倫理やプライバシーの観点、プロジェクトに影響する業界のガイドラインを管理しているか確認する
			14	分析Pで生み出された成果物の知財権の扱いを管理する(独自手法の場合は、特許申請を検討するなど)
		計画の承認	15	分析Pの実施計画について、合意形成のための調整や説明を行い、ステークホルダーの承認を得る
			16	[分析P開始] 分析業務を外部に委託する場合は、業務委託契約を結ぶ(データの取り扱い、知的財産の権利、成果物に対する責任の所在、情報開示、利用範囲など)
	組み込み後の業務設計	業務設計	17	データ分析結果を利用・適用する対象業務のプロセス等を把握/設計する
			18	取組データ別にデータの利活用及び開示のガイドラインと管理・アクセス方法をステークホルダー間で設定する
			19	対象業務の運用ルールや運用方法を決定する
			20	対象業務の目的や目標を達成し、モニタリング方針・方法やモニタリング頻度(KPI)を決定する
			21	対象業務をすでに実施している場合は、現状業務のシステム構成およびデータ仕様、業務運用を確認する
	環境整備	アーキテクチャ設計	22	対象業務を実現するためのアーキテクチャを検討し、課題解決に向けたITシステム導入のロードマップを作成する
			23	アーキテクチャ設計に基づき、分析Pに必要なクラウド環境・ハードウェア環境を整備する
24			アーキテクチャ設計に基づき、分析Pに必要な通信環境を整備する	
25			アーキテクチャ設計に基づき、分析Pに必要なソフトウェア環境を整備する	
26			不必要な情報の漏れがないように認可化を行い、事前審査や事前承認所に馴染ませる	
27	分析Pに必要な情報共有/P/管理環境を整え、必要に応じてSLAを締結する			

## タスク構造図を確認しよう

タスク構造図は、タスク分類を表す図形と、各図形をつなぐ矢印が記載されています（**図 18**）。タスク構造図には、「タスク構造図（中分類）」と「タスク構造図（小分類）」の2種類があります。データサイエンティスト業務の全体像を知りたい場合は、「タスク構造図（中分類）」を確認しましょう。データサイエンス業務の詳細なタスクの流れを知りたい場合は、「タスク構造図（小分類）」で確認できます。

**図 18** タスク構造図（中分類）



## ●タスクリストにも流れがある

### タスクの矢印をたどってみよう

タスク構造図（図 18）には、多くの矢印が記されています。実はこの矢印が、データサイエンティスト業務を理解するうえで重要なポイントです。

データサイエンティストの業務は、一つ一つのタスクを順に進めるウォーターフォール型では進行しません。タスクを行ったり来たりするため、同じタスクを何度も繰り返し実施することもあります。また、途中でうまく行かなくなった場合は、かなり前のタスクにまで戻ることもあります。

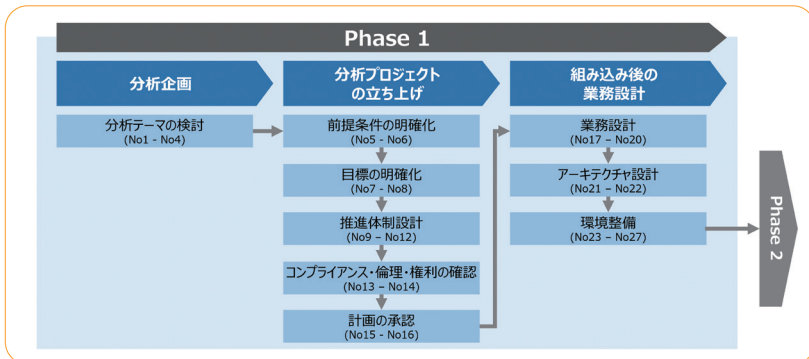
実際にタスクの流れを確認しながら、どのようにデータサイエンティスト業務が進められていくのか確認してみましょう。

#### Phase1：企画およびプロジェクトの立ち上げフェーズ

データサイエンティストの業務は「分析企画」からスタートします。取り組むべき分析テーマを決定し、分析プロジェクトを立ち上げます（図 19）。

データ分析結果を業務に組み込むための「業務設計」、「アーキテクチャ設計」、「環境整備」を行った後、次のフェーズに進みます。

図 19 Phase1 のタスクの流れ

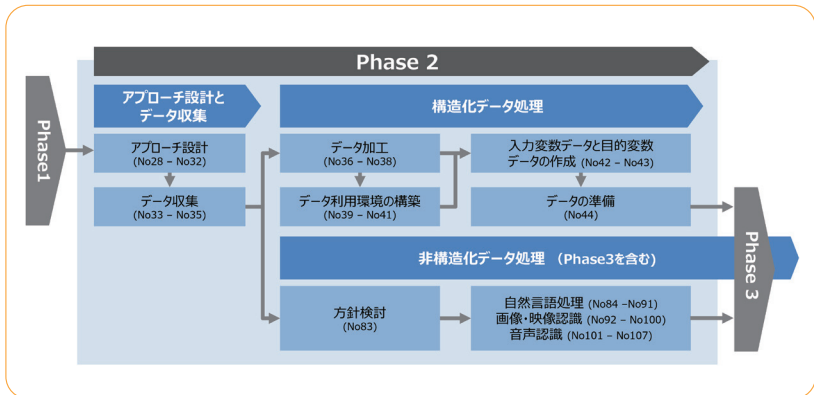


## Phase2 : アプローチ設計およびデータの収集 / 処理フェーズ

データ分析を始める前に、どのようにデータ分析を進めるのか「アプローチ設計」を行います。このアプローチ設計に基づいてデータを収集します。

次の「データ処理」に関しては、収集したデータの種類に応じてタスクが分岐します（**図 20**）。収集したデータが構造化データの場合は、「データ加工（タスク No36）」に進みます。収集したデータが、テキストや画像・映像、音声のような非構造化データの場合は、「方針検討（タスク No83）」に進みます。必要なデータ処理を終えたら、次のフェーズに進みます。

**図 20** Phase2 のタスクの流れ



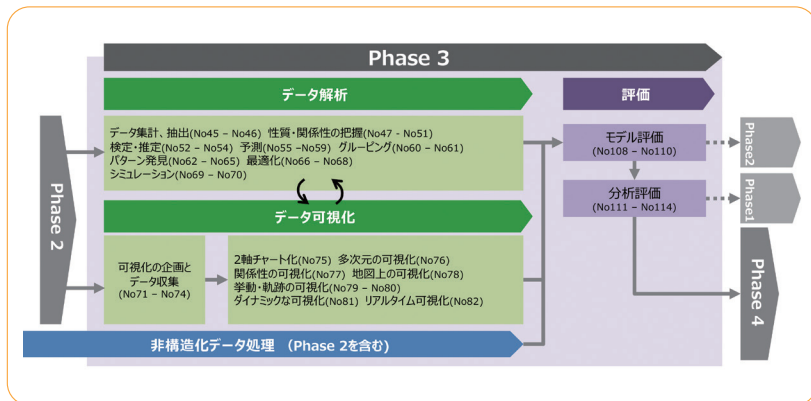
### Phase3 : データ解析およびデータ可視化フェーズ

それぞれの「データ処理」を行った後は、「データ解析」および「データ可視化」に進みます。「データ解析」タスクと「データ可視化」タスクは表裏一体で相互に行き来しながらタスクを進めるため、双方向の矢印になっています（図 21）。

非構造化データに関しては、「データ処理」と「データ解析」および「データ可視化」が密に連携しているため、「非構造化データ処理」の一部として「データ解析」、「データ可視化」を位置付けています。

データ解析・データ可視化後、「モデル評価」と「分析評価」を行います。「モデル評価」の結果が思わしくない場合は、Phase2の「アプローチ設計とデータ収集」に戻り実施タスクを見直します。また、「分析評価」の結果が思わしくない場合は、Phase1の「組み込み後の業務設計」まで戻って実施タスクを見直します。「モデル評価」および「分析評価」に問題がなければ、次のフェーズに進みます。

図 21 Phase3 のタスクの流れ

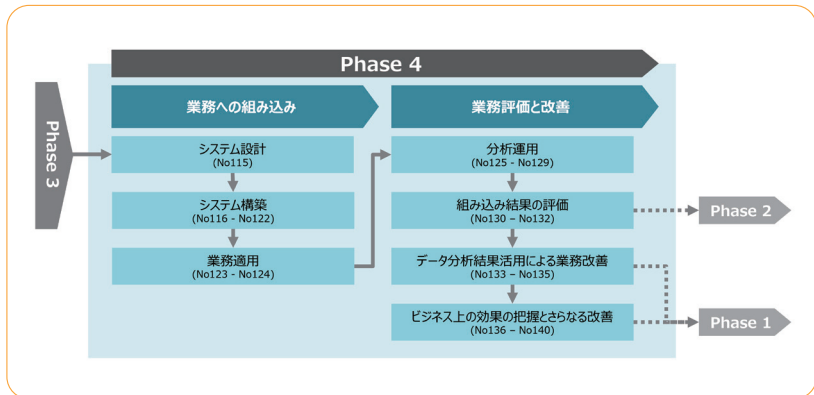


## Phase4 : 業務への組み込みおよび業務評価 / 改善フェーズ

分析結果を業務で活用するために、ソリューションや分析システムを構築します。システム構築後は分析システムを運用しながら業務成果を評価し、必要に応じて改善を行います (図 22)。

「組み込み結果の評価」が思わしくなかった場合は、Phase2 の「アプローチ設計とデータ収集」まで戻り、再度タスクを実行します。また、「データ分析結果活用による業務改善」が思わしくなかった場合は、Phase1 の「組み込み後の業務設計」まで戻り、再度タスクを実行します。

図 22 Phase4 のタスクの流れ



## すべてのタスクを実施する必要はない

タスクリストには数多くのタスクが記載されていますが、すべてを実施する必要はありません。それぞれの分析プロジェクトによって実施すべきタスクは異なるため、タスクリストを基にタスクを取捨選択します。

たとえば一部のタスクを外部に委託する場合は、「契約」に関するタスクを実施する必要があります（表5）。一方で、外部に委託しない場合は、「契約」に関するタスクを実施する必要はありません。

また、Phase3の「データ解析」や「データ可視化」に関するタスクは、分析目的に応じて適切なタスクをいくつか選択して実施します。Phase4の「業務への組み込み」においても、すでに分析システムが存在しているのであれば、システム構築の一部タスクをスキップすることができます。

このようにタスクリストは、分析プロジェクトの目的に応じて、取捨選択して使うことを覚えておきましょう。

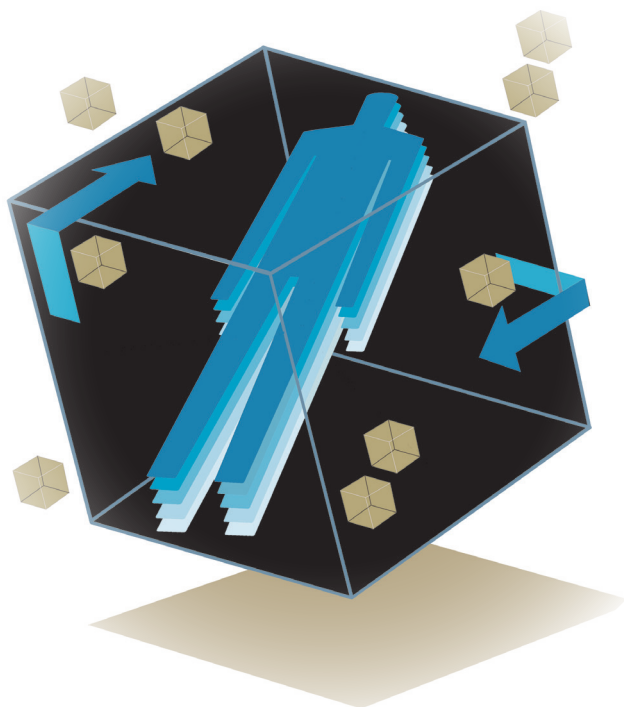
表5 契約に関するタスク

No	評価項目
4	【分析企画契約】分析企画を外部に委託する場合は、秘密保持契約および業務委託契約を結ぶ
16	【分析PJ契約】分析業務を外部に委託する場合は、業務委託契約を結ぶ（データの取り扱い、知的財産の権利、成果物に対する責任の所在、情報開示、利用範囲など）
122	【開発契約】業務への組み込みを外部に委託する場合は、システム開発のための開発契約書を結ぶ（知的財産、データの取り扱いなど）
129	【運用契約】分析システム運用やモデル見直し業務を外部に委託する場合は、分析運用のための運用契約書を結ぶ（SLAなど）



Part 05

# スキルチェックリスト・ タスクリストの使い方



## ● データサイエンティスト初心者・目指している人

データサイエンティストは、実はとても多様性のある職業の総称です。駆け出しのみなさんにとってスキルチェックリストは、データサイエンティストの全体像をつかみ、まず何から手につけるかを考えるのに役立ちます。

### データサイエンティストの全体像の把握

データサイエンティストのスキルセットは、①ビジネスカ、②データサイエンスカ、③データエンジニアリングカ——の3つの分野に大別できます。それぞれにどのような違いがあるかをみてみましょう。スキルセットごとの特徴や役割を知ることは、自分自身が目指そうとしているデータサイエンティスト像を考え、より具体的に感じられるきっかけにもなります。

### 勉強計画を立てる

いずれかの分野のスキルセットにしぼってみても、具体的なスキルは多岐にわたっています。スキルも膨大にありますが、すべてを一人で遂行しないとならないというわけではありません。

スキルチェックリストには、データサイエンス業界で使われる用語がちりばめられています。そのなかから興味があるものを探して、まず何に着手するかを考えてみるとよいでしょう。

また、Part1 で紹介したデータサイエンティスト検定を受験することで自分の理解度を測ることができ、今後の進路やキャリアの実現においてサポートとなることでしょう。

### タスクリストの使い方

タスクリストは、データサイエンティストが担う仕事の全体像を知ることができる、いわばガイドラインです。まだ自分が経験したことのない業務だとしても、タスクリストをどのスキルを学ぶべきかの目安として役立ててください。

## ● データサイエンティスト実務家

すでにデータサイエンティストとして活動している方にとっても、スキルチェックリストは、自分の守備範囲を把握し、強みを自覚しつつ、次のスキルアップへの道筋づくりに役立ちます。

### 自分のスキルのアセスメント

実務家のみなさんは、スキルチェックリストを自身のスキルのアセスメント（客観的な評価）として使うことができます。実際に自分がどのようなスキルを持っているかを確認することで、強みと弱みをはっきりと把握できます。

### スキルアップに活かす

自身のスキルの強みと弱みを知ることは、スキルアップに活かすことにつながります。たとえば、

- ・「見習い」レベルをクリアしていないスキルセットがあれば、まずその分野のスキル習得を急ぐ
- ・主軸とする分野のスキルアップを狙う
- ・全スキルセットの「★★」の必須スキルを埋める
- ・チームメイトのスキルの穴や、業務で活用できそうだがまだ持っていないスキルを伸ばす

など、さまざまな計画が考えられます。

スキルチェックリストを参考に、自分なりの計画を立てるとよいでしょう。

### タスクリストの使い方

タスクリストでは、自分が業務で担当してきた範囲や、仲間の担当範囲についてより理解を深めることができます。またプロジェクトマネージャーがどのような視点でプロジェクトを見ているかが疑似体験できるでしょう。

普段の業務における自分の担当範囲のみならず、挑戦できそうな領域があれば、そこに挑戦すべく計画を立ててみるのもいいでしょう。

## ● データサイエンティストチームリーダー ● プロジェクトマネージャー

チームやプロジェクトを率いる立場の方にとっても、スキルチェックリストとタスクリストが役に立ちます。たとえばスキルチェックリストは、チームメンバーのスキルアセスメント（スキルの客観的な評価）に活用でき、そこからプロジェクトへの配置、育成計画にも役立ちます。またタスクリストは、プロジェクトの計画策定から失敗の原因特定、人員の配置計画にも役立てることができます。

### チーム全体のスキルアセスメントと育成・採用

スキルチェックリストは、データサイエンス業務のスキルの一覧表です。メンバーにスキルのアセスメントを行ってもらうことで、今チームにどのようなスキルを持っている人がいるのかが把握できます。

またメンバーがデータサイエンティストとしてどのようなキャリアを歩みたいと考えているかを把握する際にも、スキルチェックリストの3つの分野のスキルセット（ビジネスカ、データサイエンスカ、データエンジニアリングカ）は共通言語として大いに役立ちます。

各メンバーの現状のスキルと望んでいるキャリアパスに合わせて、適切なプロジェクトへの配置、育成計画への応用、採用計画の策定にも役立てることができます。

### プロジェクト運営

タスクリストには、データサイエンティストが取り組むプロジェクトで行われるタスクが網羅されています。このなかから自身のプロジェクトに合うものを選び出せば、プロジェクトの計画策定にも大いに活用できます。

スキルチェックリストのなかには、プロジェクト運営に必要なスキルも記載されています。データサイエンティストのプロジェクトは、技術的な課題解決はもちろん、多様な立場の人が合意形成（コンセンサス）を

図りながら進めるチームワークも重要であることを忘れないようにしてください。

さらにスキルチェックリストを利用すると、誰をどの工程に配置するか、いつまでにどのようなスキルを持った人間が必要になるかなどを考えることができます。これらの事実を踏まえて、戦略的なプロジェクト運営が可能です。

また、過去の失敗や、想定より長い時間がかかったプロジェクトの原因究明にも利用できます。たとえばタスクに過不足があった、または実施の順序に問題があったということも特定できるかもしれません。

## ● 外注計画担当者

データサイエンティストの業務領域は多岐にわたっており、専門性も求められるため、一部の業務を外注するケースもあります。その際、企業が求めるスキルと候補者が提供するスキルのミスマッチが起こることもあります。

スキルチェックリストを活用すれば、こうしたミスマッチを未然に防ぐことができます。

### 必要なスキルの理解と候補者のスキルのアセスメント

データサイエンティストに外注をする際には、まず全体の工程のどの部分を外注し、どの部分を内製するかを決める必要があります。タスクリストには、今後発生するタスクが系統的に記されているので、これを参考に外注工程を決めましょう。

その次に、どのようなスキルを持った人材が必要なのかを決めることになります。スキルチェックリストから必要なものを選び出すことで、求めている人材像をはっきりさせることができます。

次に、候補者となる方のスキルのアセスメントをしましょう。候補者にスキルチェックリストを直接使ってもらったり、履歴書や面談・面接の中で使ってアセスメントしたりすることもできるでしょう。また、人材エージェントにあらかじめスキル要件を伝えることで、効率的に人材を紹介してもらうことができます。

## 処遇・待遇を決める

データサイエンティストは新しい職種であり、業務内容も多岐にわたることから、適切な評価と報酬の設定がむずかしいこともあります。プロジェクトが要求するスキル水準がわかれば、市場のなかでどの程度の待遇を用意すべきかを検討する際の参考にできます。

## ●人事採用担当者・エージェント

### 人事採用担当者の候補者理解とスキルのアセスメント

ひとことでデータサイエンティストと言っても、実際には領域は幅広く、タスクも多様です。そのため、過去の経歴でデータサイエンティストとして活躍していたとしても、自社業務で活躍できるとはかぎりません。

そこで、タスクリストを参考に自社におけるタスクの全体像を確認し、強化したいタスクを実行できるスキルを持つ人材要件を定めましょう。また、役割が違えば給与も変わるので、適切な給与と設定をするためにもスキルやタスクに基づいた職務内容の定義が必要になります。

### 転職支援エージェントにおける候補者アセスメント

求職者のスキルを、チェックリストを用いて確認します。また、求人要件の業務範囲であるタスクの経験があるかどうかを見極める際にもタスクリストが役立ちます。

## ● 学校の先生

### スキルチェックリストの使い方

対象となる学生が持つ専門性やこれまでの習熟度の違いを踏まえて、まずは「★」に絞ってスキルチェックリストを押さえていきましょう（高度な訓練や教育を受けている学生には、「★★」や「★★★」に進んでもらってください）。

社会経験が乏しい学生には、ビジネス力の項目では実感にくい部分があるかもしれません。PBL\* 型の授業などを通じてスキルを習得できるように心掛けてください。実際の課題を解決していくなかで、スキルチェックリストが持つ意味を実感できるでしょう。

スキルチェックリストがカバーする範囲はとても広いので、学生が気後れしてモチベーションを落とさないように気を配ってください。そして、スキルチェックリストを見ながら、学生自身がなりたいデータサイエンティスト像を膨らませることができるよう指導していきましょう。

### タスクリストの使い方

授業などでデータ分析を実践した際には、タスクリストを参照して現状がどのタスクに該当するかを理解してもらえるようにしましょう。

\* PBL (Project Based Learning) : 問題解決型学習。または課題解決型学習。

## ● 学生（学ぶ側の人）

### スキルチェックリストの使い方

自分自身のスキルをチェックしてみましょう。いきなり全部をチェックしようとせず、まず「★」からスタートするとよいでしょう。

ビジネス経験のない人にとって、ビジネスカの内容はハードルが高いと感じるかもしれません。「★★」や「★★★」を見ても、気後れしないでください。将来の自分はどのようなスキルを持ったデータサイエンティストでありたいか、どのような分野で活躍したいかを思い描きながら、あわてず焦らず、スキルを少しずつ身に付けていきましょう。

大学や高専の数理・データサイエンス・AI 教育で学ぶ内容（リテラシーレベル・応用基礎レベル）は、「★」のスキルに対応しています。授業の中で学んだ内容を身につけることができているか、チェックリストを見ながら確認しましょう。

### タスクリストの使い方

まずタスク構造図を眺めて、全体の流れとフィードバックループ（タスクの戻り）について理解しましょう。次にタスクリスト全体をどのようなステップでタスクが進行していくのかを考えながら見ていきましょう。

タスクリストのなかで「何を言っているのか、まったく想像つかない」という項目もあるかもしれません。しかしここで大切なのは、各フェーズの役割とタスク中分類を踏まえて目的に応じたタスクを実行することであり、タスクの流れとフィードバックループの関係を理解することです。

いきなりすべてのタスクが実行できるようにはなりません。ローマは一日してならず、着実に前に進んでいきましょう。

分析プロジェクトを担当する際にはタスクリストを参照し、現状のタスクをしっかりと把握しましょう。



## ● 育成制度を作ろう

### 重要スキルを定義する

スキルレベル列を「★」または「★★」または「★★★」でフィルタし、レベルごとに検討しましょう（図 23）。

図 23 スキルレベルのフィルタを設定する

スキルチェックリスト 2019年 改訂版 <ビジネスカ>				▼ 部分野帯りのスキル					
№	Sub N	スキルカテゴリ	スキルレベル	チェック項目	必須	必須	必須	必須	
5	5	行動規範	★★	<div style="border: 1px solid gray; padding: 5px;"> <p>並び替え</p> <p>↑ 昇順 ↓ 降順</p> <p>色別: なし</p> <p>フィルター</p> <p>色別: なし</p> <p>指定の値に等しい <input type="radio"/> ★★</p> <p><input type="radio"/> および <input type="radio"/> または</p> <p>指定の値に等しい <input type="radio"/> ★★★</p> <p>検索</p> <p><input type="checkbox"/> (すべて選択)</p> <p><input type="checkbox"/> ★</p> <p><input checked="" type="checkbox"/> ★★</p> <p><input checked="" type="checkbox"/> ★★★</p> <p>フィルターのクリア</p> </div>					
6	6	行動規範	★★						
7	7	行動規範	★★						
8	8	行動規範	★★★						
10	10	行動規範	★★						
11	11	行動規範	★★★						
13	13	行動規範	★★						
14	14	行動規範	★★						
15	15	行動規範	★★						
17	2	契約・権利保護	★★						
18	3	契約・権利保護	★★						
19	4	契約・権利保護	★★★						
20	5	契約・権利保護	★★						
21	6	契約・権利保護	★★						
21	6	契約・権利保護	★★		権利保護				

「自社必須スキル」列を作り、スキルチェックリストの定義では必須スキルになっていないものでも、自社にとっては重要なスキル項目について「○」を追加しましょう（図 24）。

スキルチェックリストのスキル項目は、特定のドメインへの依存度が高いものや個別の手法については、利用頻度が高かったり重要であったりしても、必須の「○」がっていない場合があります。

図 24 「自社必須スキル」列を作成する

スキルチェックリスト 2019年 改訂版 <ビジネス>										▼ 他分野寄りのスキル		自社必須 スキル	
№	Sub №	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	DF	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づき行動できる							○	○
2	2	行動規範	*	ビジネスマインド	データがコアプロセスや顧客体験を改善するための重要な役割を認識している								○
3	3	行動規範	*	ビジネスマインド	課題や問題を言語化することの重要性を理解している								○
4	4	行動規範	*	ビジネスマインド	現場に出向いてヒアリングするなど、一次情報と接することの重要性を理解している								○
5	5	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」を持って真早く意思決定を行うことができる								○
6	6	行動規範	**	ビジネスマインド	作業ありではなく、本質的な問題（イシュー）ありで行動できる								○
7	7	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了できる								○
8	8	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								○
9	9	行動規範	*	データ倫理	データを取り扱う人間として適切な倫理を身に付けている（データのなつ道、改ざん、盗用を行為しないなど）								○
10	10	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適切な倫理を持てるよう、適切にチームを管理できる								
11	11	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構築し、その運営を主導することができる								
12	12	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								○
13	13	行動規範	**	コンプライアンス	担当するビジネスや業界に関連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							*	
14	14	行動規範	**	コンプライアンス	個人情報の取扱いに関する法令、その他のプライバシーの問題、依頼元との契約的観点に基づき、匿名化すべきデータを識別できる（匿名化により個人を特定できるもの、依頼元がデータ別々の処理をどのように依頼し利用するのかなども考慮して）								
15	15	行動規範	**	コンプライアンス	匿名化と匿名化の違いを理解しており、適切な方法で匿名加工情報を扱うことができる							*	*
16	1	契約・権利保護	*	契約	請負契約と委任契約の違いを説明できる								

## 提供する学習方法を決める

「自社必須スキル」列を「O」でフィルタしましょう（図 25）。

図 25 「自社必須スキル」列でフィルタを設定

スキルチェックリスト 2019年 改訂版 <ビジネスカ>										▼ 他分野からのスキル	
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	自社必須スキル	▼ フィルタ	
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を試みることができる					色: なし フィルター: なし 1つ選択してください および または 1つ選択してください	
2	2	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、意味合いがない」とことを理解している					色: なし フィルター: なし 1つ選択してください および または 1つ選択してください	
3	3	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					色: なし フィルター: なし 1つ選択してください および または 1つ選択してください	
4	4	行動規範	*	ビジネスマインド	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					色: なし フィルター: なし 1つ選択してください および または 1つ選択してください	
5	5	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、時間と情報がある程度揃った状態で意思決定を行うことができる					色: なし フィルター: なし 1つ選択してください および または 1つ選択してください	
6	6	行動規範	**	ビジネスマインド	作業のスピード感だけでなく、本質的な問題（イシュー）ありきで行動できる					色: なし フィルター: なし 1つ選択してください および または 1つ選択してください	
7	7	行動規範	**	ビジネスマインド	分析で得られ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り、粘り強くタスクを完了できる					色: なし フィルター: なし 1つ選択してください および または 1つ選択してください	
8	8	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、生み出す価値観点で常に競えらるべく信頼あるアウトプットを生み出すことをコミットできる					色: なし フィルター: なし 1つ選択してください および または 1つ選択してください	
9	9	行動規範	*	データ管理	データを取り扱う人間として適切な倫理を身に付けている（データの用途を行わないなど）					色: なし フィルター: なし 1つ選択してください および または 1つ選択してください	
12	12	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、EJ-個人情報保護規程：GDPR 工務所の就業を理解し、守るべきポイントを説明できる					色: なし フィルター: なし 1つ選択してください および または 1つ選択してください	
21	6	契約・権利保護	**	権利保護	AI・モデル開発において、既存の著作権や特許の侵害を懸念・確認できる					色: なし フィルター: なし 1つ選択してください および または 1つ選択してください	
25	5	論理的思考	*	MECE	データの事象の重複に気づくことができる					色: なし フィルター: なし 1つ選択してください および または 1つ選択してください	
28	4	論理的思考	**	構造化能力	様々なデータや事象を、箇条書きやグルーピングによって、構造化できる					色: なし フィルター: なし 1つ選択してください および または 1つ選択してください	
29	5	論理的思考	*	簡潔化能力	通常見受けられる現象の場合において、分析的な意味をいかに正しく表現できる					色: なし フィルター: なし 1つ選択してください および または 1つ選択してください	
38	14	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相対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形での説明ができる					色: なし フィルター: なし 1つ選択してください および または 1つ選択してください	
39	15	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					色: なし フィルター: なし 1つ選択してください および または 1つ選択してください	
42	16	論理的思考	***	説明能力	プレゼンテーションの相手からの疑問や反論に対して、適切な応答を返すことができる					色: なし フィルター: なし 1つ選択してください および または 1つ選択してください	
52	5	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、基本的な形での説明ができる					色: なし フィルター: なし 1つ選択してください および または 1つ選択してください	
58	11	課題の定義	***	スコーピング	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、本質を抽出できる					色: なし フィルター: なし 1つ選択してください および または 1つ選択してください	

「学習方法」列を作り、学習方法を記載しましょう（図 26）。学習方法の種類としては書籍、ウェブサイト、研修、OJT などがあります。詳細な書籍名、サイト URL、研修名まで書きましょう。

図 26 「学習方法」列を作る

スキルチェックリスト 2019年 改訂版 <ビジネスカ>										▼ 他分野からのスキル		
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	自社必須スキル	学習方法	▼ フィルタ	
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を試みることができる						色: なし フィルター: なし 1つ選択してください および または 1つ選択してください	
2	2	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、意味合いがない」とことを理解している						色: なし フィルター: なし 1つ選択してください および または 1つ選択してください	
3	3	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						色: なし フィルター: なし 1つ選択してください および または 1つ選択してください	
4	4	行動規範	*	ビジネスマインド	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						色: なし フィルター: なし 1つ選択してください および または 1つ選択してください	
5	5	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、時間と情報がある程度揃った状態で意思決定を行うことができる						色: なし フィルター: なし 1つ選択してください および または 1つ選択してください	
6	6	行動規範	**	ビジネスマインド	作業のスピード感だけでなく、本質的な問題（イシュー）ありきで行動できる						色: なし フィルター: なし 1つ選択してください および または 1つ選択してください	
7	7	行動規範	**	ビジネスマインド	分析で得られ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了できる						色: なし フィルター: なし 1つ選択してください および または 1つ選択してください	
8	8	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、生み出す価値観点で常に競えらるべく信頼あるアウトプットを生み出すことをコミットできる						色: なし フィルター: なし 1つ選択してください および または 1つ選択してください	
9	9	行動規範	*	データ管理	データを取り扱う人間として適切な倫理を身に付けている（データの用途、改ざん、盗用を行わないなど）						色: なし フィルター: なし 1つ選択してください および または 1つ選択してください	
10	10	行動規範	**	データ管理	チーム全員がデータを扱う人間として適切な倫理を持つよう、適切なチームを管理できる						色: なし フィルター: なし 1つ選択してください および または 1つ選択してください	
11	11	行動規範	***	データ管理	データの取り扱いに関する、会社や関係全体の倫理を維持、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						色: なし フィルター: なし 1つ選択してください および または 1つ選択してください	
12	12	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、EJ-個人情報保護規程：GDPRなど）、必要な工務所の就業を理解し、守るべきポイントを説明できる						色: なし フィルター: なし 1つ選択してください および または 1つ選択してください	
13	13	行動規範	**	コンプライアンス	担当するビジネスや業務に関する法令を認識し、データの保持期間や運用ルールに適合することができる						色: なし フィルター: なし 1つ選択してください および または 1つ選択してください	
14	14	行動規範	**	コンプライアンス	個人情報取扱いに関する法令、その他のプライバシー（パターンの、信頼性、信頼性）の契約的観点に基づき、匿名化すべきデータを識別できる（匿名化により個人を特定できるもの、仮名化がデータ処理の結果をどのように保持・利用するのにも考慮して）						色: なし フィルター: なし 1つ選択してください および または 1つ選択してください	

## 確認方法を定める

育成施策は多くの場合、成果として育成者数の報告を求められます。必要に応じて「自社必須スキル」を満たした人数がカウントできるように、「確認方法」を定めておきましょう。

図 27 確認方法を記載する

A		E		F		G		H		I		J		K		L			
スキル訂版 <ビジネスカ>																			
▼ 他分野育りのスキル																			
NO	サブカテゴリ	チェック項目				DE	DS	必須スキル	自社必須スキル	学習方法	確認方法								
1	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							○	○									
2	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							○	○									
3	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							○	○									
4	ビジネスマインド	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							○	○									
5	ビジネスマインド	現場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」を持って集中し意思決定を行うことができる							○	○									
6	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							○	○									
7	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了できる							○	○									
8	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							○	○									
9	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）							○	○									
10	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																	
11	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																	
12	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							○	○									

確認方法については、以下を参考にしてください。

- 多くのスキルに該当する研修は、研修受講を確認の条件とする
- 書籍やウェブサイトが該当するスキルは、知識確認のテストを用意する
- OJTしか該当しないスキルは、経験を確認する記述式の設問を用意する
- 可能であれば「★」は研修とテストのみが望ましい
- 「★★」以降は経験を確認する必要性が高く、読む人（審査員）を必要とする
- 「★★」以降は研修受講を振るい落としの条件として使える
- 「★★」以降は、知識確認よりも、ケーススタディ、ハッカソン、コンテスト、アクションラーニング等での成果物で評価する

Part 06

# リストにない定義や考え方



## ●ドメインスキル

### 汎用性の高いスキルを中心に定義

---

データサイエンティストが関わる業務は、多岐にわたります。データ活用の方向性を決めたり分析結果を読み解いたりするには、その業務に特有の知識が必要不可欠になります。しかし本スキルチェックリストでは、特定の業務で求められるスキルについては取り上げていません。本スキルチェックリストでは、なるべく特定のドメイン（業務の領域）に寄らない、汎用性の高いスキルを中心に定義しています。

たとえば、デジタルマーケティングに携わるデータサイエンティストにとってマーケティングの知識は必要不可欠ですが、品質管理に携わるデータサイエンティストにマーケティングの知識は必要ありません。逆に品質管理に携わるデータサイエンティストには、QC（Quality Control）の知識が要求されますが、デジタルマーケティングに携わるデータサイエンティストには、QCの知識は不要です。

データサイエンティストはあらゆる事業ドメインにおいて活躍しており、今後もさらなる活躍が期待されています。そのため、ドメイン知識を漏らさずスキルとして定義することは現実的には不可能に近く、また一般性のあるスキル定義も困難と考えられることから、スキルチェックリストでは定義していません。

スキルチェックリストを活用する際は、対象ドメイン固有のスキルを加えると、より有効に活用できます。

## ● 文章に込められた意味

### ビジネスカ

---

ビジネスマインドに関するすべてのスキル項目は、必須となっています。ここでは、データサイエンティストは価値創出にコミットし、実際に価値を生み出せる人材でなければならないという意志が込められています。

このほか、No.6のスキル項目での「ザックリ感」という表現は、時間と情報が限られているビジネスの現場において、迅速かつ最善の意思決定をして前へ進んでいくという姿勢を示しています。「仮説検証思考」、「イシュードリブン」といった思考・姿勢と関連付けて理解していただきたいスキル項目です。

また、スキルカテゴリにある「着想・デザイン」のデザインとは、製品のデザインといった造形ではなく、「系」のデザインを意味します。予測モデルが勝手に価値を生み出すことはなく、予測モデルが業務フローのどの個所でどのような役割を果たすのか、全体の関連をデザインすることの重要性を表現しています。

### データサイエンスカ

---

データサイエンスを実践するにあたっては、「知識として身に付いている」こと、そして「実装できる」ことが必要です。

「説明できる」という言葉で表現されているスキルの場合、「★」では知識として「説明できる」レベルであることを意味します。これに対して「★★★」における「説明できる」は、「実装できる」レベルに達しているスキルと知識が身に付いていることを意味します。

たとえばデータサイエンスカ No.184 では、「★」レベルとして『教師あり学習』『教師なし学習』の違いを理解している」というスキルを求めています。さらに、「★★」レベルになると「教師あり学習と強化学習の違いを、前提、定義、応用先といった観点で説明できる」となり、実装に

必要なスキルが求められます。

また、No.207「不均衡データ (Imbalanced Data) がモデルに与える影響を理解し、サンプリングや評価／損失関数のチューニングなどを適切に対処できる」のように、「★★」レベルでは、特定の条件下において必要となるスキルが求められます。

## データエンジニアリング力

---

データエンジニアリング力では、「設計できる」「構築できる」といった表現を多用しています。

スキル項目に記載されている内容と同一の経験や実績がなくても、作業を行うために必要なスキルや知識があり、設計や構築の手順を説明できる状態であれば「○」と評価してよいでしょう。

データエンジニアリング力のスキル項目では、記載する技術やアーキテクチャを具体的にイメージしてもらうために、比較的知名度の高い製品名やソフトウェア名を例示していることがあります。これらは理解促進のための例示であり、取り上げている製品の利用を推奨する意図はありません。

また、環境構築のスキル項目では、一般的なオフィス用 PC 1 台で処理できるもの、専用サーバが必要になるもの、分散処理が必要となるもので、レベルに違いが出てきます。これを表すために、データ量の目安として、「数十万レコードのデータ」、「数千万レコードのデータ」、「数十億レコードのデータ」といった数字を例示しています。ただし実際には、業務の要件やデータの特性によって環境構築の判断基準となるデータ量は変化するので、注意が必要です。

データエンジニアリングの分野において、一般的なシステムエンジニアとデータサイエンティストのスキルに違いはあるかということを目にします。実際、システムエンジニアとデータサイエンティストの間では、多くのスキルが重複します。両者の違いは、システムエンジニアにとっては与えられた要件を元に高品質なシステムを実装する設計力・開発力が重視されるのに対し、データサイエンティストにとっては大規模で複合的なデータ分析環境構築に必要な要件を定義するスキルが重視されることです。

また、自然言語処理や画像・映像認識、音声認識などに代表されるよう



な高度で専門性が高いエンジニアには、データサイエンティストとしての基礎的な部分を押さえることでプロジェクト全体を俯瞰し、チームとしての開発力を確実なものにする役割が求められます。

つまりデータサイエンティストには、より広い知識と上流の視点でシステムを俯瞰する能力が求められるのです。

## タスク

---

タスクリスト Phase3 の「評価」タスクは、モデル評価 (No.108 ~ No.110) と分析評価 (No.111 ~ No.114) に分けています。

データ解析の評価というと、モデルの精度を評価する「モデル評価」を想像しがちですが、一連のデータサイエンティスト業務のなかでは、「モデル評価」と「分析評価」の両方を実施する必要があります。

活動目的に対する有効性・妥当性を判断する「分析評価」を忘れてほしくないため、明示的に「評価」を二つの項目に分けてタスクを定義しています。

## ● 進化の速い分野

データサイエンティストに必要な技術・知識の広がりや進化・深化の速さに対応するため、スキルチェックリストやタスクリストの改訂のたびに、すべての分野において、スキル項目の追加・削除やレベルを精査・見直しをしています。

## ビジネスカ

---

ビジネスカのなかでも、マインド、論理思考、課題定義といった本質的な問題解決での思考力に関する部分は、変化が少ないスキルカテゴリです。

一方で、法令・契約・権利保護に関しては、常に最新のルールを理解しておくことが求められます。たとえば、GDPR（General Data Protection Regulation：一般データ保護規制）が、2018年5月25日施行されたことを受けて、2019年版のスキルチェックリストでは、新たに関連するスキル項目を記載しました。

また、2021年版では、AI活用に関わる倫理スキル（No.11, No.12）や着想・デザインスキル（No.45～51など）を新規に追加し、ビジネス観点でのAI活用スキルを大きく変更しました。

法令・契約・権利保護については、常に最新のルールを捉えておく姿勢が重要であると共に、倫理的、実装的な側面でのAIを扱うスキルの重要性が増しています。

## データサイエンスカ

---

データサイエンスの分野では、日々新たな技術や理論が誕生し実践されています。ほんの数年前までは理論の提唱と一部先進企業による実践にとどまっていた技術や理論も、活用に広がりを見せています。例えば、2019年版のスキルチェックリストで追加されたディープラーニング技術の高度化による生成モデルの発展（No.206）について、今日ではフェイ

ク画像やフェイクニュースなどに関する話題を目にする機会が多くなったかと思います。自然言語処理モデルについても様々な企業で実践が進んでおり、2021年版では新たに Attention 機構に関する理解を追加しています (No.221)。また、機械学習の活用が進む中、予測モデルを故意に誤らせるためのデータを加える敵対的サンプルといった脅威も認識されており、このようなリスクに対する理解も 2021 年版ではあらたなスキル項目として追加されています (No.204)。

既存の技術や理論においても、データ活用が広がる中で活用・実践に進展が見られます。施策などの介入効果において、比較するグループ間での効果に対する影響の偏り (セレクションバイアス) をできる限り排除し、介入効果をより正確に推定するための因果推論などはその一例でしょう。従来、医療や教育の効果分析などを中心に研究や実践が進んでいましたが、今日では広告出稿などランダム化比較試験が難しいマーケティング施策の効果推定においても、様々な企業で実践されています。このような流れを受け、2021 年版では因果推論についてサブカテゴリを新設しました。

また、データ活用の広がりにより技術・理論の普及や新たなライブラリの登場などが進むことで、既存技術の難易度は変化していきます。ロジスティック回帰や最尤法 (No.65) に関する知識などは判別問題で広く使われるようになり、2019 年版のスキルチェックリストでは「★★★」から「★★」にスキルレベルが変化しました。また、ロジスティック回帰ができることの理解 (No.60) など、どの手法で何ができるかの理解に関しては、2021 年版でレベルを落としているものがあります。

2019 年版で追加された AI の出力結果に対する公平性 (No.209) のような知識も、AI が社会に浸透していくなかで重要性はますます高まっています。

データサイエンスカのスキルを高めていくには、スキルチェックリストの内容だけでなく、最新の技術動向や活用動向を学んでいくことが重要です。

## データエンジニアリング力

---

近年はクラウドベンダーの提供する分析サービスが拡充され、大量の

データを安価に処理できるようになってきました。そのため、多数のサービスメニューから要件に合うクラウドサービスを選択して分析環境を構築するスキルが重要になりつつあります。

また、以前は複雑だった GPU (Graphics Processing Unit) の制御についても、扱いやすいライブラリが登場しています。これにより、GPU 制御そのものの理解よりもライブラリの特徴を捉えて有効に活用するスキルの方が重要になりつつあり、それに応じたスキル項目を追加しています。さらに、ライブラリの普及で実装が容易になったスキル項目については、スキルレベルを見直しています。

2019 年版のスキルチェックリストでは、ブロックチェーンや量子コンピュータなど、まだ発展途上にある新しい技術に関するスキル項目を追加し、2021 年版では、クラウド活用をインフラ構築の標準スキルとし、AI の実用化を前提に、AI システム運用に関するスキル項目も追加しました。

データサイエンティストは、新しい技術動向には常に目を配り、将来そうした最新技術が普及段階に入った際に、それを採用するか否かを見極める眼力・能力が求められます。

## タスク

---

タスクは、本質的な流れは変わらないため、細分化されることはあっても骨格はあまり変化しません。一方で、タスクリスト Phase3 のデータ解析、データ可視化、非構造化データ処理で参考例として記載している各種活用事例については、日々新しい事例が登場しているため、常に最新の情報を収集する姿勢が重要です。



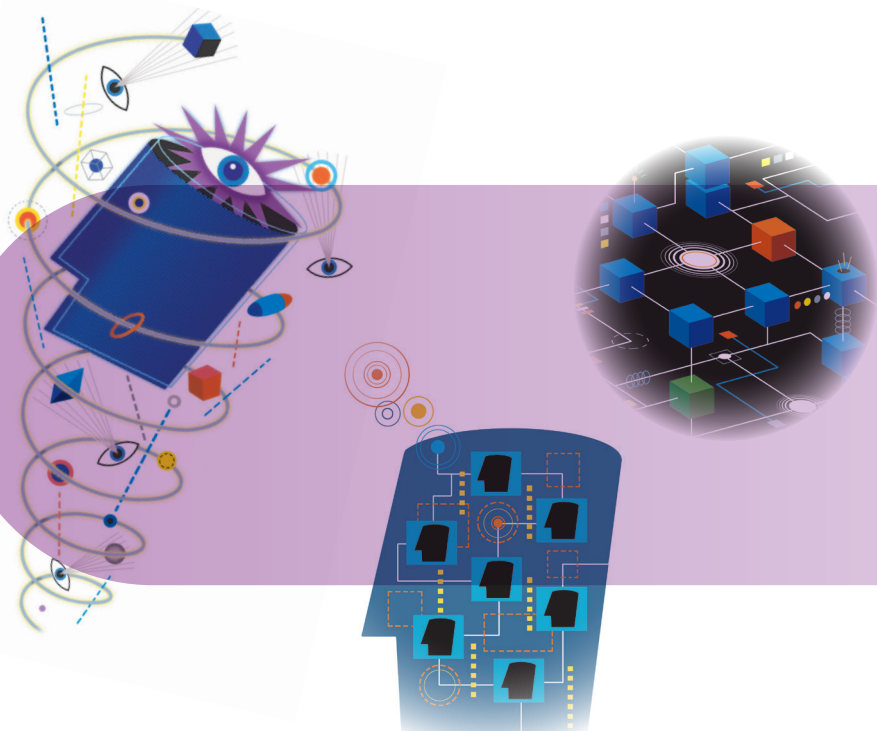
## ● ビジネスや事業への関わり方

### データサイエンティストはビジネスパーソンたれ！

現在の日本社会では、データドリブンな経営推進や事業推進、デジタルトランスフォーメーションのためのデータ活用など、データサイエンティストが活躍できる場が飛躍的に広がっています。

その一方で、データサイエンティストが事業に加わっても、なかなか結果が出ないということもよく聞かれます。なぜなのでしょう？

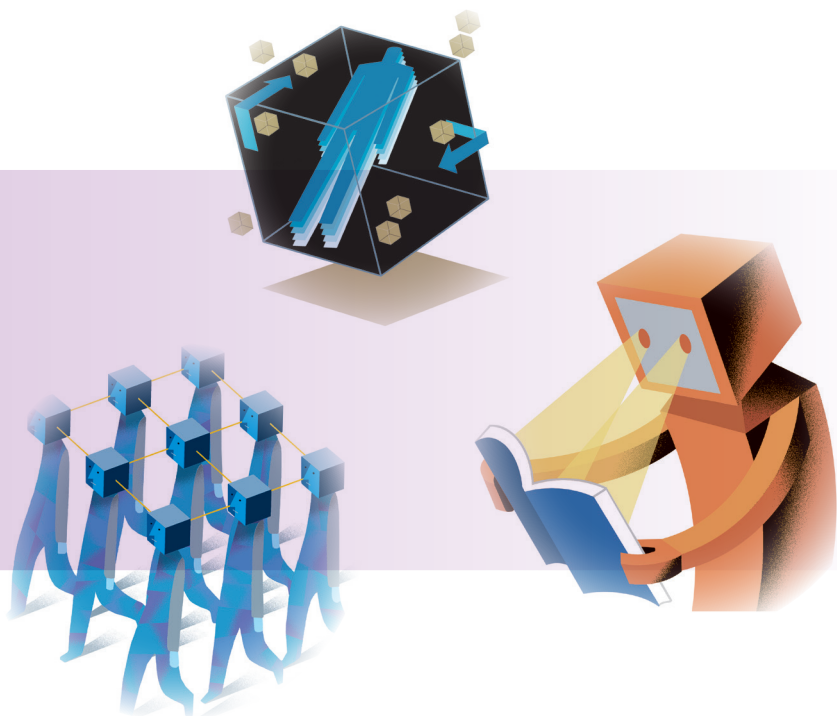
ビジネスシーンにおいては、「実践力」や「現場力」が重んじられることがよくあります。データサイエンティストはもちろん科学的に、論理的に物事を捉え、データの価値を最大化する役割を担っているわけですが、「現場」ではなかなか理解を得られにくいものです。



「現場」や「事業」、あるいは「経営」を背負っている人、チーム、組織のメンバーとの相互理解を深め、協働し、データのカで新しい事業価値を生み出すために、データサイエンティストは「現場」にもっと積極的に入り込むしかないので。

それにはいろいろな方法があるでしょう。企業風土や組織の立ち位置によっても変わるかもしれませんが。たとえば、現場で通用するくらいの「現場実践力」を高めていくということもあります。それを「言語化」し、「ロジカルに説明」すること。情熱を持って数字や分析結果を武器にして、現場同士をネットワークし導くこと。現場や事業の中期目標に合わせ、目利き力を生かして先回りのデータ活用戦略を立てること……これらを体得することは、これからのビジネスパーソンにとって当たり前のスキルなのかもしれません。

ビジネスカにもそのエッセンスは含まれています。データサイエンティストは、常にビジネスパーソンでもありたいものです。



## ● お役立ち情報

一般社団法人データサイエンティスト協会では、さまざまなお役立ち情報を公開しています。ぜひご活用ください。

### ▼ データサイエンティスト検定 リテラシーレベル

2021年9月からデータサイエンティスト協会が始まった検定試験です。スキルチェックリストで定義しているアシスタント・データサイエンティスト（見習いレベル:★）と数理・データサイエンス教育強化拠点コンソーシアムが公開している数理・データサイエンス・AI（リテラシーレベル）におけるモデルカリキュラムを総合し、実務能力と知識を有することを証明する試験となっています。データサイエンティスト初学者やこれから目指す学生、社会人の方を対象にしています。

毎年、春と秋に実施し、参考図書や対策講座も出ています。学んだことや実践していることの腕試しにもなりますが、これを学ぶことで実務的な能力を身につけ、社会の役に立てるデータサイエンティストを目指すきっかけとなることでしょうか。詳細は協会 HP に記載されています。

<https://www.datascientist.or.jp/dskentei/>

### ▼ データサイエンス 100 本ノック（構造化データ加工編）

構造化データについて、無料で利用できるオープンデータが多数ありますが、データ分析力の実践という観点で環境整備がなされたものは、あまり多くありません。

分析実務においては構造化データの活用が多くを占めるという実態を鑑み、データと実行環境構築スクリプト、そして演習問題をワンセットにして GitHub にて公開しています。

<https://github.com/The-Japan-DataScientist-Society/100knocks-preprocess>

### ▼ スキルチェックリスト（GitHub 版）

スキルチェックリストのさらなる普及を目的として、csv 版や検索を行



えるビューワーを GitHub にて公開しています。

・ **csv 版スキルチェックリスト**

<https://github.com/The-Japan-DataScientist-Society/skills-checklist>

・ **スキルチェックリストビューワー**

<https://the-japan-datascientist-society.github.io/skills-checklist-viewer/>

## 一般社団法人データサイエンティスト協会

一般社団法人データサイエンティスト協会は、新しい職種であるデータサイエンティストに必要なスキル・知識を定義し、育成支援や評価制度の構築など、高度 IT 人材の育成と業界の健全な発展への貢献、啓蒙活動を行っています。

また、所属を超えてデータ分析に関わる人材が開かれた環境で交流や議論をして自由に情報共有したり、意見を発信できる場を提供しています。

<https://www.datascientist.or.jp/>

### データサイエンティストのための スキルチェックリスト / タスクリスト概説 Skill checklist & Task list Overview

2020年7月7日 初版発行

2022年4月1日 第二版発行

監 修：一般社団法人 データサイエンティスト協会

発 行：独立行政法人 情報処理推進機構（IPA）

社会基盤センター人材プラットフォーム部

〒113-6591 東京都文京区本駒込二丁目28番8号

文京グリーンコートセンターオフィス

e-mail：ikc-sx@ipa.go.jp

W e b：https://www.ipa.go.jp

制作 / デザインレイアウト：ACMEX INC. デザイン室

イラストレーション：ネモト円筆



監修：一般社団法人データサイエンティスト協会



監修 / 発行：独立行政法人情報処理推進機構（IPA）