

2017年10月25日

プレスリリース
報道関係各位

一般社団法人データサイエンティスト協会

**「データサイエンティスト スキルチェックリスト ver.2」および
「データサイエンス領域タスクリスト」を発表**

このたび、データサイエンティスト協会 スキル委員会（委員長：安宅 和人、副委員長：佐伯 諭）は、10月23日（月）に開催した「データサイエンティスト協会 4th シンポジウム」内において発表した、データサイエンティストに必要とされるスキルをまとめた「データサイエンティスト スキルチェックリスト」の第2版を公開いたしました。

本内容は、2014年～2015年に公開したデータサイエンティストの「ミッション、スキルセット、定義、スキルレベル」および「スキルチェックリスト」を、現在のビジネス環境や実態にあわせて全面的に見直したものです。改定のポイントは大きく以下の通りです。

- 機械学習、深層学習時代に対応した項目の見直し・追加
- 「画像・動画処理」「音声/音楽処理」のカテゴリ・項目見直し（データサイエンス領域）
- 「知財」カテゴリをビジネス領域に追加
- 「データの理解・検証」「分析プロセス」をビジネス領域からデータサイエンス領域に移動
- 各スキルレベルでの必須スキル項目を全面的に見直し、増加（全項目の約1/4を指定）
- 各領域内のカテゴリと項目の見直しを行い、項目数も422項目から457項目に増加
- 以上をふまえ、データサイエンティストのスキルレベルを総合的に説明した内容を見直し

また、本年4月には大量データを分析し、その分析結果を活用するための一連のデータサイエンティスト業務を整理した「データサイエンス領域 タスクリスト」の第1版を、独立行政法人情報処理推進機構（IPA）が定める「ITSS+（プラス）」として公開したことを、あわせてお知らせいたします。

本タスクリストは、各スキル要件がどのようなプロセスに必要とされるかを整理した、横断的な内容で、ITスキル標準（ITSS）の新領域「データサイエンス領域」にて採用された「スキルチェックリスト」を補完するものとなっています。プロセスの全体観と必要なスキル項目がセットで明確化されることを通じ、より一層、習得、スキルの利活用に向けた道筋が明確化することを期待しております。



スキル委員会は、2014年にデータサイエンティストの「ミッション、スキルセット、定義、スキルレベル」、2015年に「データサイエンティスト スキルチェックリスト」、2016年に422のスキル項目がインターネット上でセルフチェックできるウェブサイト(*1)を立ち上げるなど、データサイエンティストという新たな職種が健全に成長をするための環境整備に取り組んでまいりました。

なお、セルフチェックができるウェブサイトは今回の改定に合わせ、本年12月に更新する予定です。

- ・「スキルチェックリスト ver2」は、以下よりダウンロードいただけます。

URL : https://www.slideshare.net/DataScientist_JP/2017-81179087

- ・「データサイエンティスト協会 4th シンポジウム」での「スキル委員会発表資料」は、以下よりダウンロードいただけます。

URL : https://www.slideshare.net/DataScientist_JP/4th-81178634

- ・IPAから発表されたデータサイエンス領域のタスクリストは、以下よりダウンロードいただけます。

URL : <https://www.ipa.go.jp/jinzai/itss/itssplus.html>

(*1) http://www.datascientist.or.jp/dssjournal/ds_skillcheck/

<ご参考情報>

【データサイエンティストのスキルレベル】

	ビジネス力 (business problem solving)	データサイエンス力 (data science)	データエンジニアリング力 (data engineering)
Senior Data Scientist 業界代表 レベル	<ul style="list-style-type: none"> ・業界を代表するデータプロフェッショナルとして、組織全体や市場全体レベルでのインパクトを生み出すことができる ・対象とする事業全体、産業領域における課題の切り分け、論点の明確化・構造化 ・新たなデータ分析、解析、利活用領域の開拓 ・組織・会社・産業を横断したデータコンソーシアムの構築、推進 ・事業や産業全体に対するデータ分析を核としたバリューチェーン創出など 	<ul style="list-style-type: none"> ・業界を代表するデータプロフェッショナルとして、データサイエンスにおける既存手法の限界を打ち破り、新たに課題解決可能な領域を切り拓いている ・既存手法では対応困難な課題に対する新規の分析アプローチの開発・実践・横展開 ・高難度の分析プロジェクトのアプローチ設計、推進、完遂能力など 	<ul style="list-style-type: none"> ・業界を代表するアーキテクトとして、データサイエンス領域で行おうとしている分析アプローチを、挑戦的な課題であっても安定的に実現できる ・複数のデータソースを統合した例外的規模のデータシステム、もしくはデータプロダクトの構築、全体最適化 ・技術的限界を熟知し、これまでにない提案の提示・実行など
Full Data Scientist 棟梁 レベル	<ul style="list-style-type: none"> ・生み出す価値にコミットするプロフェッショナルとして、データサイエンティストとは何かを体現したビジネス判断、課題解決ができる ・初見の事業領域に向かい合う場合や、スコープが複数の事業にまたがる場合であっても本質的な課題を見出し、構造化・深掘りができる ・解決に必要な結果を総合した上で、説得力ある形で共有し、関連する組織、人を動かし、知見の横展開、組織を超えるつなぎ込みができる ・プロフェッショナルからなる複数のチームによるプロジェクトの役割、目標を定義、推進し、全体としてのアウトプットにコミットできると共に、メンバーを育成、さらには持続的な育成システムを作り出すことができる 	<ul style="list-style-type: none"> ・予測、グルーピング、機械学習、大量データの可視化、言語処理、最適化問題などの応用的なデータサイエンス関連のスキルを活かし、データ分析プロジェクトの技能的軸を担うことができる ・初見の事業領域に向かい合う場合や、スコープが複数の事業にまたがる場合であっても、適切な分析・解析アプローチの設計、実行、深掘りができる ・複数もしくは高度な分析プロジェクトを持つチームにおいて、Associate Data Scientist（独り立ちレベル）以下のメンバーの技能を育成することができる 	<ul style="list-style-type: none"> ・数十億レコード程度の分析環境の要件定義・設計、データ収集/蓄積/加工/共有プロセスや IT セキュリティに関するデータエンジニアリング関連のスキルを活かし、データ分析プロジェクトを中核的に推進することができる ・全体を統括するアーキテクトとして、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総合的に把握し、設計や開発に活かすことができる ・複数もしくは高度な分析プロジェクトを持つチームにおいて、Associate Data Scientist（独り立ちレベル）以下のメンバーの技能を育成することができる



	ビジネス力 (business problem solving)	データサイエンス力 (data science)	データエンジニアリング力 (data engineering)
Associate Data Scientist 独り立ち レベル	<ul style="list-style-type: none"> 大半のケースで自立したプロフェッショナルとして、ビジネス判断、課題解決ができる -ビジネス要件の整理、プロジェクトの企画・提案 -知財リスクの確認などの適切な対応 既知の領域、テーマであれば、新規課題であっても解くべき問題の見極めや構造化、深掘りができる データ、分析結果に対する表面的な意味合いを超えた洞察力を持ち、担当プロジェクトの検討結果を取りまとめ、現場への説明、実装を自律的かつ論理的に行うことができる 5名前後のプロフェッショナルによるチームでのプロジェクトを推進しアウトプットにコミットできる -タスクの粘り強い完遂 -イシュードリブンでスピード感のある判断 -プロジェクトマネジメントと個別メンバーの育成 -異なるスキル分野の専門家、事業者との協働など 	<ul style="list-style-type: none"> 単一プロジェクトにおけるデータ分析を Full Data Scientist (棟梁レベル) に相談しつつ推進できる Assistant Data Scientist (見習いレベル) の日々の活動に適切な指示ができる 既知の領域、テーマであれば、新規課題であっても適切な分析・解析アプローチの設計、実行、深掘りができる 基礎的なデータ加工については、自律的に実施できる -外れ値・異常値・欠損値の対応 -適切な学習データとテストデータの作成 基礎的な分析活動については、自律的に実施できる -多重共線性を考慮した重回帰分析 -パラメトリックな 2 群の検定の活用 (t 検定) -適切な初期値設定を行った非階層クラスター分析 -主成分分析や因子分析 -機械学習における過学習の理解 -形態素解析などを用いた基本的文書構造解析など 	<ul style="list-style-type: none"> 単一プロジェクトにおけるデータ処理・環境構築を Full Data Scientist (棟梁レベル) に相談しつつ推進できる Assistant Data Scientist (見習いレベル) の日々の活動に適切な指示ができる 数千万レコード程度のデータ処理・環境構築については自律的に実施できる -データの重要性や分析要件に則したシステム要件定義 -適切なデータフロー図、論理データモデル作成 -Hadoop や Spark での管理対象データ選定 -SDK や API、ライブラリなどの適切な活用 -SQL の構文理解と実行 -分析プログラムのロジック理解と分析結果検証など 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 データ匿名化方法の理解と加工処理の設計ができる
Assistant Data Scientist 見習い レベル	<ul style="list-style-type: none"> ビジネスにおける論理とデータの重要性を理解したデータプロフェッショナルとして行動規範と判断が身についている -データを取り扱う倫理と法令の理解 -引き受けたことは逃げずにやり切るコミットメント -迅速な報告や、報告に対する指摘のすみやかな理解など データドリブンな分析的アプローチの基本が身につけており、仮説や既知の問題が与えられた中で、必要なデータを手入れし、分析、取りまとめることができる -データや事象のダブリとモレの判断力 -分析前の目的、ゴール設定 -目的に即したデータ入手 -分析結果の意味合いの正しい言語化 -モニタリングの重要性理解など 担当する検討領域についての基本的な課題の枠組みを理解できる -担当する業界の主要な変数 (KPI) -基本的なビジネスフレームワークなど 	<ul style="list-style-type: none"> 統計数理の基礎知識を有している (代表値、分散、標準偏差、正規分布、条件付き確率、母集団、相関、ベイズの定理など) データ分析の基礎知識を有している -予測 (回帰係数、標準誤差・) -検定 (帰無仮説、対立仮説・) -グルーピング (教師あり学習、教師なし学習・) 適切な指示のもとに、データ加工を実施できる -基本統計量や分布の確認、および前処理 (外れ値・異常値・欠損値の除去・変換や標準化など) データ可視化の基礎知識を有している (ヒストグラム、散布図、積み上げ棒グラフなど) 	<ul style="list-style-type: none"> データやデータベースに関する基礎知識を有している -構造化/非構造化データの判別、論理モデル作成 -ER 図やテーブル定義書の理解 -SDK や API の概要理解など 数十万件程度のデータ加工技術を有している -ソート、結合、集計、フィルタリングができる -SQL で簡単な SELECT 文を記述・実行できる -設計書に基づき、プログラム実装できる 適切な指示のもとに、以下を実施できる -同種のデータを統合するシステム的设计 -インポート、レコード挿入、エクスポート セキュリティの基礎知識を有している (機密性、可用性、完全性の 3 要素など)



	ビジネス力 (business problem solving)	データサイエンス力 (data science)	データエンジニアリング力 (data engineering)
DS 以前の方	<ul style="list-style-type: none">・ビジネスは勘と経験だけで回すものだと思っている・課題を解決する際に、そもそも定量化する意識がない	<ul style="list-style-type: none">・基本統計量の意味を正しく理解していない・指数を指数で割り算したりする・「平均年収」をそのまま鵜呑みにしたりする・グラフ・チャートの使い方が不適切	<ul style="list-style-type: none">・レポートされてくる数値サマリに目は通すが、特に記憶には残らない・アクセス解析システムを使っていない・Excel や Access は数字しか入れない

【スキルカテゴリー一覧】

スキルカテゴリー一覧							
カテゴリ		サブカテゴリ	項目数	カテゴリ		サブカテゴリ	項目数
データサイエンス 力	1	統計数理基礎	16	データ エンジニアリン グ力	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		データエンジニアリング力 項目数		129
	10	データの理解・検証	23	ビジネス力	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		ビジネス力項目数		100
データサイエンス力項目数			228	スキル項目数合計			457

●一般社団法人データサイエンティスト協会について <http://www.datascientist.or.jp/>

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となるスキル・知識を定義し、育成のカリキュラム作成、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、自由に情報共有や意見発信ができる場を提供しています。2017年10月現在、77社8団体の法人会員と約7,960名の一般（個人）会員が参画しています。代表理事：草野 隆史（株式会社ブレインパッド 代表取締役社長）、所在地：東京都港区、設立：2013年5月。

●報道関係の方からのお問い合わせ

一般社団法人データサイエンティスト協会事務局

TEL：03-6721-9001、e-mail：info@datascientist.or.jp

*本プレスリリースに記載されている会社名・商品名は、それぞれの権利者の商標または登録商標です。

*本プレスリリースに掲載されている情報は、発表日現在の情報です。

以上