

KADOKAWAグループにおける データの活用事例の紹介と 求めるデータサイエンティスト像について

(株)KADOKAWA Connected 塚本 圭一郎

自己紹介

塚本圭一郎

- 略歴

- 2014-03 大阪大学大学院 情報科学研究科 博士後期課程修了
- 2014-04 (株)ドワンゴ 入社
- 2019-04 (株) KADOKAWA Connected 兼務出向
- 2020-04 (株) KADOKAWA 兼務出向

- 所属

- (株)ドワンゴ ICTサービス本部 Dwango DataManagement Service部 部長
- (株) KADOKAWA Connected Integrated Data Service部 部長
- (株) KADOKAWA DX戦略アーキテクト局 Integrated Data Service部 部長

- 職務内容

- KADOKAWAグループにおける部署横断的なデータ活用促進をデータマネジメントからリード



目次

- 会社紹介
- 弊社事例の紹介
- 企業の求めるデータサイエンティストとは
- まとめ

会社紹介

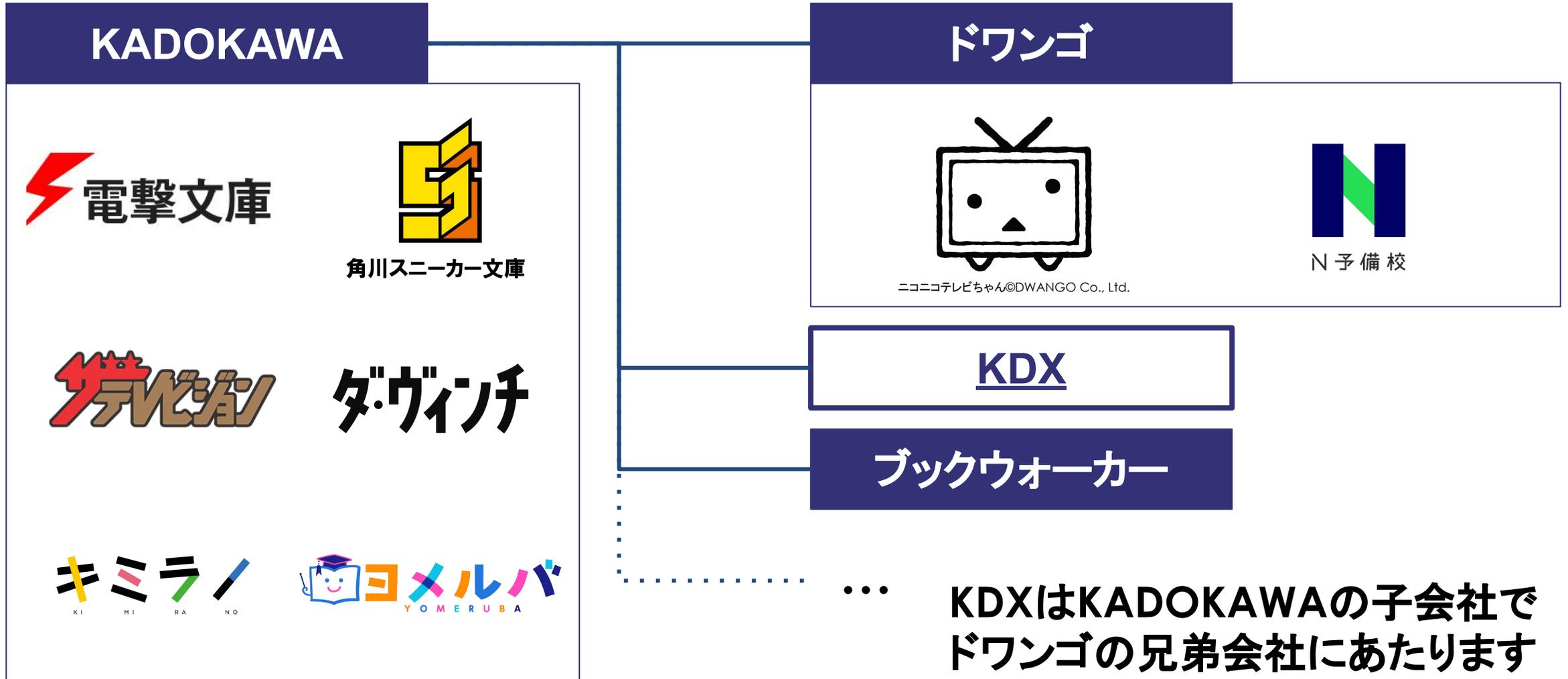
KADOKAWA Connected (KDX) とは

日本の文化をベースに働く人々の「生涯生産性」を最高に高めるためのソリューションを提供する企業

事業概要

- KADOKAWAグループへのICTサービス提供
 - niconico事業へのICT基盤提供
 - KADOKAWA系IPのWebサイト基盤提供
- 働き方改革の実現支援
 - Office向けICTインフラサービス提供
 - 導入後も含めたコンサルティングサービス

KADOKAWAグループとは



事例紹介

niconico事業のビッグデータ分析基盤の提供

ID発行数	月間UB	ログインMAU	プレミアム会員	チャンネル入会
約8,017万	約5,747万	約646万	約161万	約123万

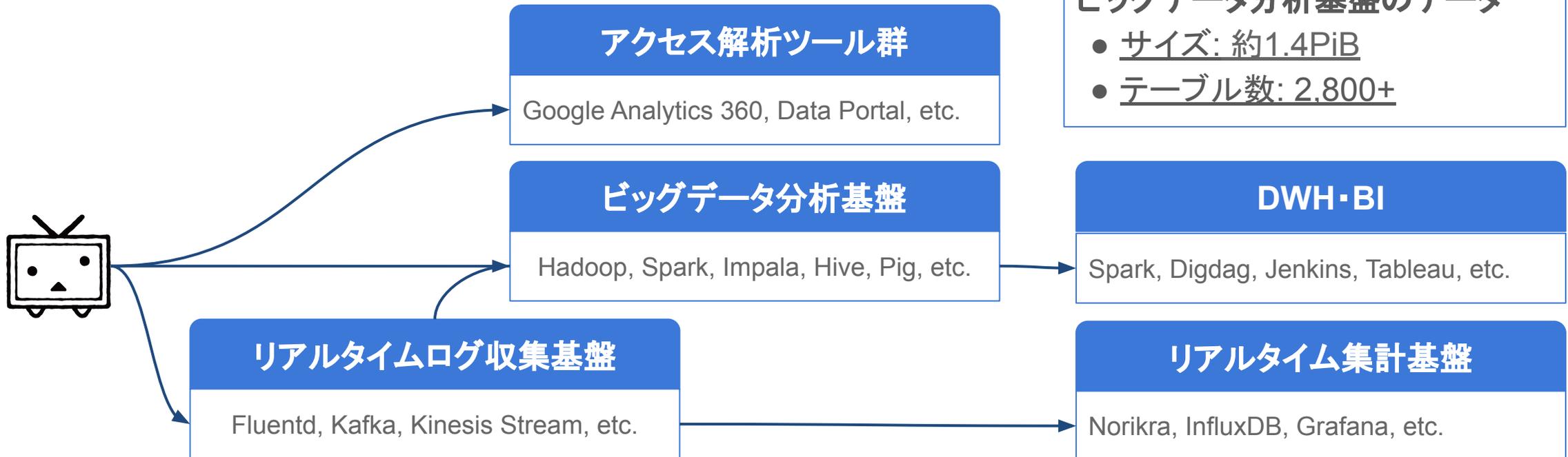
2020年6月末時点

分析者数

- アクセス解析ツール群: 410+
- ビッグデータ分析基盤: 260+

ビッグデータ分析基盤のデータ

- サイズ: 約1.4PiB
- テーブル数: 2,800+

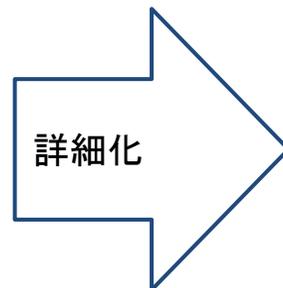


管理会計ツールの開発と保守

事業単位よりも詳細なサービス単位の損益計算書などを作成するツールを提供、より細やかな意思決定を支援

niconico事業

売上	1,000,000,000
人件費	300,000,000
インフラ費	400,000,000
...	...
合計	200,000,000



ニコニコ動画

売上	400,000,000
人件費	100,000,000
インフラ費	100,000,000
...	...
合計	60,000,000

ニコニコ生放送

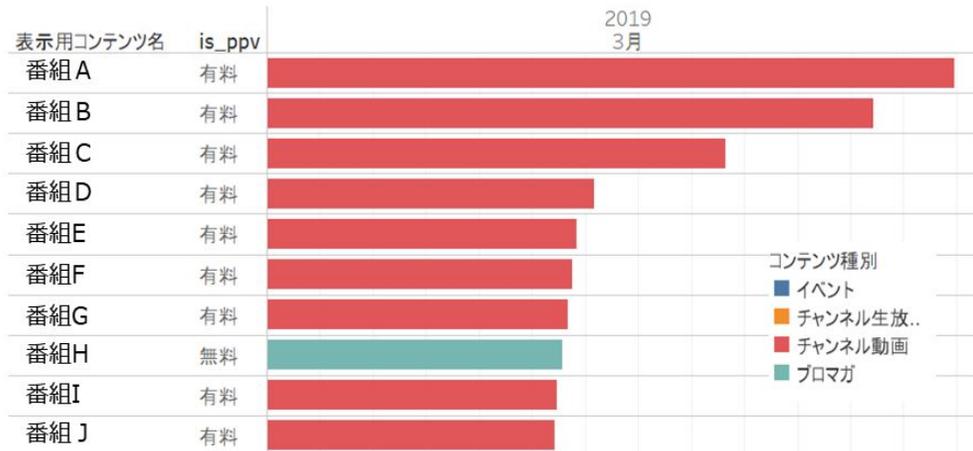
売上	400,000,000
人件費	80,000,000
インフラ費	150,000,000
...	...
合計	40,000,000

数値は勿論、仮値です

チャンネルオーナー向けダッシュボードの提供

チャンネル事業のコンサルチームがチャンネルオーナーにデータに基づく提案が可能になるダッシュボードを開発

コンテンツ横断で最も入会に寄与した
コンテンツのランキング表示

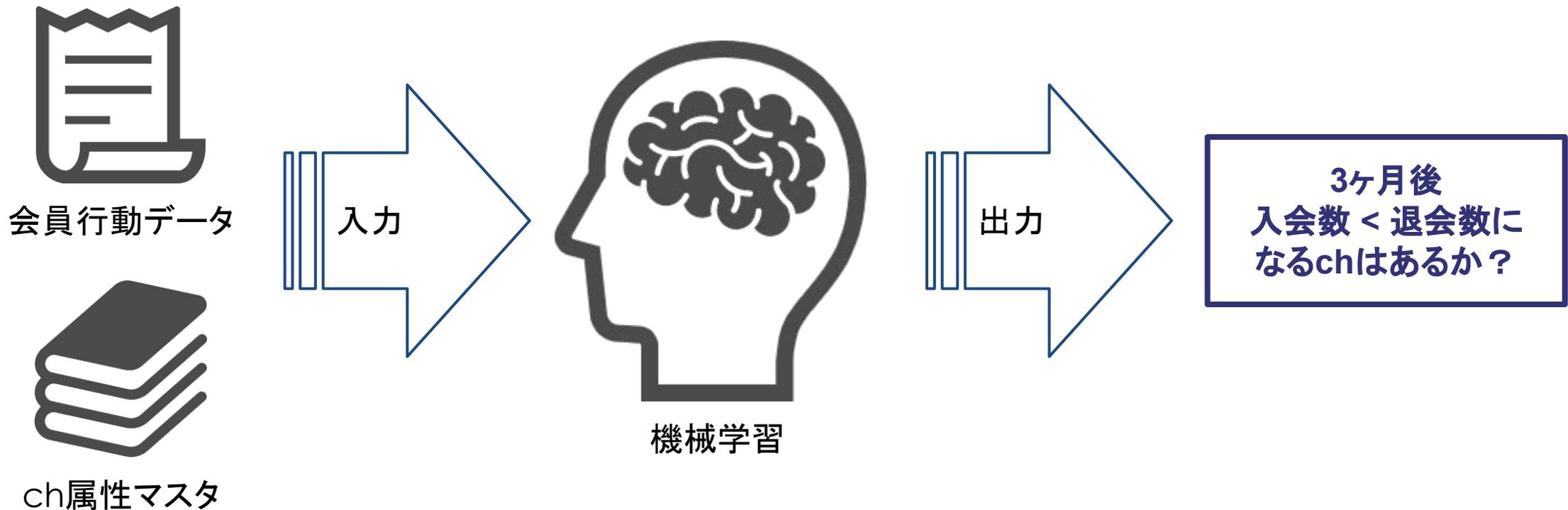


新規コンテンツ/過去コンテンツが
どの割合で入会に寄与しているか



純減チャンネル予測モデルの構築 (PoC段階)

将来不調になるチャンネルを予測し、コンサルチームが優先的に対応しなければならないオーナーを明らかにする



GPUクラスタ基盤の提供 (予定)

ドワンゴのMLエンジニアリングチームに紅莉栖に代わる強力なGPUクラスタ基盤を提供する

ニュース

ドワンゴ、ディープラーニング専用GPUサーバー「紅莉栖(くりす)」を構築
～TITAN XのQuad SLIを100台。人工知能研究用に無償貸し出し

(2015/9/17 18:13)

ツイート リスト B! 74 Pocket 108 いいね! 0 シェア



【参考画像1】 歌澤紅莉栖
©2008-2015 MAGES/Spb/Nitropius

【参考画像2】 右:クリスティーナ 左:おまいり

株式会社ドワンゴは17日、ディープラーニング専用GPUサーバーファーム「紅莉栖(くりす)」を構築し、一部の研究機関を対象に人工知能研究用として無償貸し出しを行なうと発表した。

(出典)<https://pc.watch.impress.co.jp/docs/news/721741.html>

MLエンジニアリングチームは、レコメンドシステムから将棋AIや麻雀AIまで幅広い機械学習プロジェクトを推進するチームです

詳細は、<https://dmv.nico/ja/>



(出典)<https://twitter.com/NAGA025/> (麻雀AIのナーガ)

現行DL専用GPUサーバー紅莉栖を
新世代のGPUクラスタ基盤に更改

企業が求めるデータサイエンティストとは

企業はデータサイエンティストを求めている

KPI管理や効果測定 (記述的) が企業で当たり前になり、
今後は診断的・予測的アナリティクスのニーズが高まる

原因の分析と将来の予測がこれからは必要

- 診断的: 多変量分析、RCT、集積分析など
- 予測的: 回帰分析、機械学習など

実際に事業で機械学習を用いる試みが当たり前前の時代になりつつある

(参考) <https://www.sbbbit.jp/article/cont1/36820>

データサイエンスはコモディティ化した

データサイエンス単体の希少価値が失われつつある

- 大学やオンライン学習でのデータサイエンス教育の一般化
- 様々な研究分野で機械学習が利用される状況
- Auto MLの台頭
 - DataRobot, AutoML Tables, ...

(参考) <https://www.itmedia.co.jp/enterprise/articles/2007/30/news033.html>

データサイエンス×○○○が必要になっている

× エンジニアリング

- MLOps
- データエンジニアリング
 - Volume: ビッグデータ
 - Velocity: ストリーミングデータ
 - Variety: 非構造化データ

× R&D

- アルゴリズム開発
- 未踏領域への適用
 - 新しいXXXTech

× ビジネス

- アナリティクスプラットフォーム
- 説明可能なAI
 - Explainable AI

ただし、データソースの壁がある

企業側しかこれらを学ぶために潤沢なデータや実例を持っていない
⇒ 産学連携？ or [データ流通促進](#)？

まとめ

まとめ

- 企業活動においてデータ活用は当たり前ものになり、予測的アナリティクスの需要も高まっている
- 需要に対して、データサイエンティストを大量に育成した結果、データサイエンスの希少価値が減少している
- これからは、データサイエンスに加えて異分野の専門性をもつT型人材になることが求められている

We're Hiring!

エンジニアリングで、
日本の文化を支える人たちが、
楽しい、働きやすい環境を。



KADOKAWA Connected 採用

検索

ご清聴ありがとうございました