

グローバル企業が求める データサイエンティスト

小林 哲郎

データサイエンティストという職種が注目を浴びようになり、近年その活動内容や人物像に非常に注目が集まっている。その一方で、データサイエンティスト発祥の地であるアメリカにおけるデータサイエンティストの業務内容の実態について、日本に紹介したものはほとんど見られない。本稿では、アメリカカリフォルニア州の IT 企業である Pivotal のデータサイエンスチームを例に、その職場風景、採用プロセス、業務内容、そして情報共有方法について紹介することで、グローバル企業におけるデータサイエンティストの実態を解説する。

キーワード：データサイエンティスト、グローバル企業

1. はじめに

データサイエンティストという職種が注目を浴びようになり、近年その活動内容や人物像に非常に注目が集まっている。ビッグデータやファストデータといった世の中のデータ分析に対するニーズが高まっていることもあり、多くのメディアでデータサイエンティストについての特集が組まれてきた。日本におけるこのデータサイエンティストのある種の流行はアメリカ合衆国（以下、アメリカ）に端を発しており、アメリカでは日本に先駆けてデータサイエンティストが企業で活躍している姿を多くみかける。また教育の現場においても、University of Virginia [1] をはじめとして複数の大学でデータサイエンス専門の学位を取得できる学科が開設されており、アメリカはデータサイエンスに関する取組みにおいて世界の一步先を行っていることは間違いない。

このようにデータサイエンスの最先端に行くアメリカのデータサイエンスの実態、データサイエンティストの日々の業務等について、日本に紹介されている事例はほとんどない。何人かの世界的に有名なデータサイエンティストが紹介されたり、日本人が個人的にブログ等で彼らの体験に基づいて紹介しているものが断片的には存在しているが、アメリカ等海外のデータサイエンティストたちが実際にどのような業務を行っているのかについては全容が見えてこない。そこで本稿

では、弊社 Pivotal を事例に、アメリカ発のグローバル企業におけるデータサイエンティストの職場風景、採用プロセス、業務内容、そして情報共有方法について紹介する。その際、アメリカカリフォルニア州の Pivotal 本社にて 2014 年 3 月 10 日から 22 日にかけて開催された新入社員向けデータサイエンティスト研修での私の体験を交えて紹介させていただく。なお、Pivotal は 2013 年 4 月からスタートした、カリフォルニア州パロアルト市に本拠地を置く IT 企業であり、社内データサイエンティストが世界数拠点に約 50 名在籍しており、データサイエンスを事業の大きな柱の一つとしている。

2. Pivotal 本社の雰囲気

2014 年 3 月 10 日午前 8 時半。前日借りたばかりの慣れないレンタカーを運転し、これまた慣れない道をカーナビだけを頼りに走らせ、カリフォルニア州サンフランシスコ郊外、パロアルト市内の Pivotal 本社に到着した。車から出ると少し肌寒い。3 月のカリフォルニア州パロアルトは湿度が少々高めで温暖な気候だが、1 日の気温差が大きい。車を出たところで、たまたま同じタイミングで本社に到着したシンガポールのデータサイエンティストと出会い、握手を交わし一緒に本社建物へ向かった。

本社の建物を見たときに思ったのは、「至ってシンプルだ」ということだけであった。2 階建てで外壁が白く、何も特筆することはない建物だ。後で聞いたことだが、シリコンバレーにある企業は、これと同じようなシンプルな建物を本社オフィスにしていることが多いそうだ。入り口付近に Pivotal 自転車が停まっている

こばやし てつお
株式会社 Pivotal ジャパン
〒151-0053 東京都渋谷区代々木 2-1-1 新宿マインズタワー内



写真1



写真2

たので、物珍しさに思わず立ち止まって本社の入り口の写真を1枚撮った(写真1)。本社に入ると、シンガポールの彼は慣れた振る舞いで入館手続きを済ませていた。聞くところによると、二度目のパロアルト訪問だということだ。私はその様子を見ながら、私の分まで手続きを依頼し、彼の後について研修ルームへ向かった。研修ルームにはすでに何人かのデータサイエンティストが先に到着していたので、席に着く前にまずは挨拶をした。アメリカ国内の参加者も何名かいたが、それ以外はイギリス、ドイツ、インド、イスラエル、中国、オーストラリアから、総勢20名の参加であった。いろいろな訛りのある英語を聞きながら握手と「Nice to meet you」を交わし、とりあえず自分の席を確保した。少しメールチェックでもしようかと思った矢先、先ほどのシンガポールの彼が、「ごはん食べた?」と聞いてくる。時差ボケもあって明け方に少し仮眠を取れただけだったため、朝食を摂る時間を取れていなかった。「まだだけど、なぜ?」と聞くと、「じゃ、食べに行こう」と彼は私を手招きして、研修ルームの外へ行ってしまった。

研修ルームの外はカフェテリアスペース兼キッチンスペースになっていることに気がついた。いくつかのテーブルの上に、おいしそうなお料理が並んでいるのにも気づいた(写真2)。たぶん、目よりも鼻のほうが先に料理を感知したと思う。シンガポールの彼から「ここではブレックファーストは毎日タダで提供されているんだ」と聞かされ、早速料理を取り、近くにいたデータサイエンティストたちと歓談しながらブレックファーストを楽しんだ。周りを観察していると、出勤してくる社員の多くはまずこのカフェテリアに寄って朝食を摂ってから自分のデスクに行くようだ。お決まりの「Good morning」や「How have you been?」といったフレーズで会話が始まり、簡単に近況報告をしたり、各々が聞



写真3

わっているプロジェクトについて情報交換をしたりしていた。このブレックファーストは会社からのサービスではあるが、始業前の社員同士がコミュニケーションを行う場として重要な機能を持っているのである。ほかに、キッチンスペースには、コンロのような加熱器具はないものの、水道、冷蔵庫、スナックの棚が並んでいる(写真3)。ここにある食べ物や飲み物はすべて無料であり、好きな時に好きなだけ飲食が可能である。私がおここに常駐すれば必ず太ると確信した。日本採用であったことに感謝した瞬間だった。

3. Pivotal データサイエンスチームの多様性

ブレックファースト時間が終了し、午前9時から研修がスタートした。新人研修は5日間で、毎日午前9時から午後6時の日程で行われた。その主な内容は、Pivotal データサイエンスの概要、求められる業務内容の説明、Pivotal が扱う製品群とデータサイエンスツールの紹介およびその実習であった。個々のトピックで入れ替わり立ち替わり本社在籍のデータサイエンティストが登壇し、プレゼンテーションを行った。R、Pythonなどのプログラミング言語を駆使したアルゴリズムの



写真 4

開発方法を紹介する者、Pivotal が実際に関わったデータ分析事例とその実績について紹介する者、プロジェクトマネージャーとしてのデータサイエンティストの業務内容を紹介する者、MADlib（後述のPivotal 開発の分析ツール群）等最新の分析テクノロジーについて紹介する者など、非常に充実した講師陣であった。特に、データ分析事例の紹介は、それぞれの顧客を実際に担当したデータサイエンティスト本人から話を聞く機会を持つことができたので、実際のデータ分析現場でのトラブルやプロジェクト成功の秘訣など、数多くのエッセンスを聞くことができ、非常に有意義であった。また、紹介された分析ツールについては実習も用意されており、学んだことをすぐに実践する機会が与えられたことも参加者の満足度を高めた。

登壇するデータサイエンティストたちを見ていて感じたのは、彼らのバックグラウンドが非常に多様だということだ。Pivotal には約 50 名のデータサイエンティストが在籍しているが、それぞれの出身地、専門知識、Pivotal に入社したきっかけなどはばらばらであり、一貫していないところが面白い。出身地を見てみると、インド出身者が比較的多く、次いでアメリカ、イギリス、中国、ドイツの出身者が続き、ほかにシンガポール、台湾、トルコ、オーストラリア、韓国、日本出身者が在籍しており、国際色豊かだ（写真 4）。また、女性データサイエンティストの数も少なくない。特に、ヘルスケア関連のデータ分析担当チームに在籍する 5 名のうち、3 名が女性である。

Pivotal データサイエンティストの専門領域も多彩であり、チームとして幅広い知識とスキルを持ち合わせている。彼らの修士号あるいは博士号の学位を聞いて

みると、コンピューターサイエンス、数学、統計学、オペレーションズリサーチ出身のメンバーが複数在籍しており、いかにもデータサイエンスという学位で納得する。しかし一方で、物理学、機械工学、生物学、化学、地理学（私）など、IT に必ずしも直結しないような分野のメンバーも多くいることがわかった。この多様性があるからこそ、Pivotal はさまざまなデータサイエンスの課題への解決策を提供できるのである。ある顧客を担当するデータサイエンティストがその分野に明るくない場合も多々あるが、50 名の誰かはその分野について詳しいという場合がほとんどであるため、データサイエンティスト間で密にコミュニケーションを取ることで、顧客にとって最適なデータ分析手法を提供することが可能となる。Pivotal データサイエンスサービスの提供実績領域は、医療、製造、小売、公共サービス、金融、サイバーセキュリティ、オンラインマーケティングなど多様である。

以上のような幅広い専門性を持つ Pivotal データサイエンスチームは、複数のプロジェクトチームから構成されており、世界の複数拠点から業務を遂行している。本社のあるパロアルトオフィスには 2 チームが常駐しており、1 つはサイバーセキュリティ担当チーム、もう 1 チームは製造業や小売業、公共サービスなど幅広い業界のデータ分析を担当するチームである。少し離れたサンフランシスコ市内のオフィスにはヘルスケア担当チームが在籍している。また、アメリカ国内では数名が在宅勤務の形態で雇用されており、それぞれがチーム配属されている。各チームは、経験・実績豊富なプリンシパルデータサイエンティストという上級職のデータサイエンティストが統括しており、チームメンバーはシニアデータサイエンティスト、そしてデータサイエンティストという 2 種類のデータサイエンティストメンバーで構成されている。また、（アメリカから見た）海外拠点はイギリス、ドイツ、シンガポール、日本、オーストラリアにあり、それぞれにデータサイエンティストが在籍している。

4. Pivotal のデータサイエンティスト採用プロセス

以上のように多彩 Pivotal データサイエンスのメンバーだが、彼らが Pivotal に入社する経緯もまた多様である。研修に来ていたデータサイエンティスト全員に話を聞いたところ、自ら Pivotal を見つけて応募したというケースはほとんどないことがわかった。では、どのようなルートで Pivotal を知るようになったのか。

ほとんどの場合が人材紹介会社のリクルーターから連絡を受け、初めて Pivotal の存在を知ったようである。特に多いのが、ほかの IT 関連企業からヘッドハンティングされたケース、学生時代にリクルーターから連絡を受けたケースであることがわかった。また数は少ないが、ソーシャルメディアの LinkedIn を通じて Pivotal 入社に至ったメンバーもあり、海外においてはソーシャルメディアを活用した就職活動や転職活動も活発です。に当たり前になっていることも実感できた。

Pivotal にデータサイエンティストとして入社したいという意思表示をすると、当然だが採用試験が待っている。リクルーターから連絡をもらったからといって、同時に採用通知が送られてくるわけではない。Pivotal のデータサイエンティスト採用試験では、Pivotal のデータサイエンティストとしての資質を網羅的に評価される。その資質とは、一般的に言われるデータサイエンティストの必須スキルと言われるものと似ているが、1) 業務知識や学術的な専門知識の有無、2) コミュニケーションスキル、3) IT のスキル、4) 数理・統計の知識、5) プログラミングスキル、6) 英語の能力の 6 種類である。

上記の資質の評価は面接とデータ分析テストの 2 種類に分けられる。まずは面接試験を受けることになるが、これまで携わってきた主な学術的な研究あるいはビジネスにおける実務経験についての説明を求められる。その際に、これまで利用していたデータベースの種類、プログラミング言語や分析モデル、分析結果の説明など、分析スキルや IT に関する知識を問われる。会話をしているので、コミュニケーションや英語のスキルも同時に試されることになる。また、面接官から口頭でデータ分析課題を与えられ、その課題に対してどのような回答をするか、という質問も設定されている。素早い発想力を試されている瞬間である。面接ではコミュニケーションスキルや英語のスキルの評価に重点が置かれており、その他のデータ分析関連スキルは後述のデータ分析テストで評価されることになる。なお、Pivotal のデータサイエンスチームは世界数カ所に拠点があるため、基本的に面接はスカイプあるいは電話で行われる。私の場合は、シンガポールやアメリカと日本をつないでのスカイプ面接がほとんどであった。

複数回の面接試験を突破した候補者には、最後の採用試験としてデータ分析テストが与えられ、データ分析のスキルが総合的に評価される。このデータ分析テストが課される理由は、面接する側の考えとして、「応募者がデータ分析に関する質問について、口頭でどれ

だけ上手に答えられたとしても、実際にデータを分析させてみなければ分析スキルは客観的に評価できない」、という考えに基づいている。テストを受ける側は、一般に公開されている統計データを分析し、そこから得られた有益な情報をデータ分析レポートという形で提出する。制限時間は 24 時間である。これ以上テストの具体的な内容をここで明かせないのが残念だが、その内容で興味深いのは、課題に対する答えが一意に特定されていない点である。特定の答えを導くためのテストという体裁になっていないため、提出されるデータ分析レポートの内容はバラエティに富むことになる。このようなテストを課す理由としては、「データ分析には初めから決まった答えが用意されているわけではない場合がほとんどである」、という Pivotal データサイエンスの考えが根底にある。実際のビジネスの現場におけるデータ分析で結果を出すためには、与えられたデータを用いて、限られた時間内に顧客のビジネスにとって価値あるデータ分析結果を提供することが求められる。そこには初めから答えが用意されているのではなく、分析者の思考パターンや論理パターン、これまでの経験、感性等が非常に重要となる。このデータ分析テストでは、擬似的にビジネスの現場を体験し、分析者がどのように対処するのかを審査しているのである。

さらに、提出する分析レポートには分析のプロセスをすべて記載することが求められている。つまり、一般的な分析プロセスである、データのクレンジング、変数選択、モデル選択、モデル作成と利用する分析ツール、分析結果の評価と今後の課題、という一連の分析の過程において行った作業行程をすべて試されていることになる。後日、ほかの候補者たちが提出した分析レポートを見る機会があったが、分析者によってその手法や解釈は十人十色で、このテストによって分析者の特徴や分析能力、いわば分析の「クセ」のようなものが見て取れると感じた。

5. Pivotal データサイエンティストの業務内容

新人データサイエンティスト研修が終了した後の 1 週間は、個々のデータサイエンティストとの個別ミーティングや、データサイエンスチームの全体ミーティング、実案件への参加などに時間を費やすことができた。その 1 週間バロアルトとサンフランシスコのオフィスに滞在することで、Pivotal データサイエンティストの日々の業務内容を観察することができた。その業務内

容を大別すると、1) データ分析関連業務、2) データサイエンストレーニング、3) データ分析ツールの研究開発、4) チーム内の知識やスキルの情報共有の4項目になる。以下に、それぞれについて詳しく紹介する。

5.1 データ分析関連業務

データサイエンティストの核となるデータ分析およびその関連業務は、顧客のニーズに合わせて2種類用意されている。1つは顧客のデータ活用の構想立案を行うもので、PivotalではLab Primer（ラブプライマー）と呼ばれるものである。このPrimerという言葉には「点火薬」という意味があり、「データ分析プロジェクトに火をつける」、あるいは「データ分析プロジェクトを立ち上げる」という意味が込められている。このLab Primerは、これからデータ分析プロジェクトを立ち上げたいという企業あるいは団体向けに提供するものであり、その目的は、各企業・団体に特化したデータ分析テーマの列挙とそれらの優先順位付けを行い、長期的なデータ分析プロジェクトの計画案を作成することである。具体的な内容としては、前半は企業の現状分析やブレインストーミングによるテーマの列挙を行い、後半は列挙された各テーマについて、主にビジネス上の効果、実現可能性の2点により評価が行われ、テーマの優先順位付けが行われる。この際、上記の2点のできるだけ高いものを優先的に選択することを目標とする。優先順位上位に位置づけられた分析テーマについては、具体的な分析計画案を作成し、最終成果物とする。また、この構想立案の中には実際に分析作業を行う社内の人材育成についても指針がまとめられる。この一連の作業は顧客とのディスカッションによって行われるので、担当するPivotalデータサイエンティストには非常に高いコミュニケーション能力が求められる。顧客のビジネスの実態を理解し、そのビジネスに貢献するデータ分析課題を特定することが、PivotalデータサイエンティストのLab Primerにおけるミッションである。

もう1つのサービスが、実践的なデータ分析プロジェクトの支援である。データ活用の構想を確立し優先的に分析すべきテーマが決定した後は、実際にデータ分析を行う作業が待っている。顧客のニーズによってその規模や期間が異なるが、分析テーマを1つ選択し、4~12週間の期間で1つのデータ分析のサイクルを経験することに重点が置かれる。ニーズとしては、企業のサンプルデータを用いて試験的に分析アルゴリズムを作ってみたいというもの、実際にビジネスにおける収益増加やコスト削減につながる成果を出すための分

析アルゴリズムを開発したいというもの、人材育成のトレーニングの一環として実データの分析を経験することでOJT的に分析作業を行いたいというもの、などさまざまである。どの場合においても、基本的に複数名のPivotalデータサイエンティストがプロジェクトに従事することになっている。複数名が関わる理由としては、異なる専門性の融合による相乗効果への期待、分析に対する考え方のバイアスを極力軽減する効果、そして時間の無駄を防ぐ効果が挙げられる。専門性が異なる複数のデータサイエンティストが1つのプロジェクトに関わることで、分析のアイデアが何倍にも膨らみ、1人で分析アイデアを考えた場合に比べてより良い結果を生み出せるようになる。1人で分析を担当すると、分析のアプローチにどうしてもバイアスがかかってしまうことが多く、思うような分析結果を出せない場合が少なからずあり、結果として時間を無駄にしてしまう。複数名の知識やスキルを融合させることによるプラスの効果を最大化し、マイナスな効果を軽減させることが、2名1チームでのプロジェクト従事を行う理由である。

5.2 データサイエンストレーニング

企業におけるデータ分析プロジェクトの大きな柱の1つである人材育成という課題に対し、トレーニングプログラムの開発、提供を行うこともPivotalデータサイエンティストの重要な業務である。このデータサイエンストレーニングは2014年6月時点で2種類のトレーニングが定期的に提供されている。1つは基礎的なデータ分析の理論および実習を提供するData Science and Big Data Analytics [2]である。これは5日間の日程で行われるトレーニングコースであり、データ分析の進め方、統計学、機械学習やデータマイニングの手法の理論を学び、さらに実習においてRやSQLによるデータ分析を体験するという、盛りだくさんの内容となっている。そしてもう1つのトレーニングは、ビジネスにおけるビッグデータ分析の価値や社内ビッグデータプロジェクト立ち上げの指針など、ビジネスリーダー向けのData Science and Big Data Analytics for Business Transformation [3]である。こちらも講義だけではなく、グループ実習がトレーニングに組み込まれており、ある（架空の）企業のケーススタディを基に実際のデータ分析プロジェクトを進めるための意思決定を擬似的に体験することができる。さらに主要なデータ分析手法とその活用例も学ぶことができ、1日間という短い時間の中でビッグデータ分析プロジェクトを進めるために必要なリソースや作業などを網羅的

に学ぶことができる。Pivotal データサイエンスチームでは現在、より高度な分析技術を学ぶためのコースを開発中である。そのコースの実習には、Hadoop 環境上でのデータマイニング、ソーシャルメディアデータを含むテキストマイニング、グラフマイニングといった、近年注目されている高度なデータ分析演習が数多く含まれている。

5.3 データ分析手法の研究開発

データサイエンティストという職種は主にビジネスの世界に限った活動をしていると思われがちだが、Pivotal データサイエンティストはアカデミックな活動も積極的に行っている。それがデータ分析手法の研究開発業務であり、1) 新たなデータ分析手法の開発と、2) プログラミングによる新規アルゴリズムの実装作業の 2 つに大別される。Pivotal データサイエンティストは日々企業・団体からの依頼に応え、前述のトレーニングやデータ分析課題に取り組んでいる。その活動の中で培った経験から、新たなデータ分析手法の開発を行うメンバーが少なからず在籍している。たとえば、バロアルトのサイバーセキュリティ分析チームは Knowledge Discovery in Databases の分野 [4] で、シンガポールの分析チームメンバーの一人は人口知能関連分野 [5] で、それぞれ査読付き論文を発表している。特にシンガポールのメンバーが執筆した論文は、出版された Journal of Artificial Intelligence Research より、その影響力の高さから 2014 年のベスト論文賞を受賞した [6]。また、分析手法の特許出願も積極的に進めており、セキュリティ分析の異常検知手法やヘルスケア分野における画像解析処理の手法で現在特許出願中である。

Pivotal が積極的に関わるもう 1 つの研究開発業務として、新規アルゴリズムの実装作業がある。具体的には、オープンソースのデータ分析ライブラリである MADLib の機能拡張をする作業を行っている。MADlib は 2010 年に始まったカリフォルニア大学バークレー校と Pivotal の前進である株式会社 EMC の Gremlin 事業部との共同プロジェクトで開発された SQL ベースの機械学習アルゴリズムライブラリである。その後、スタンフォード大学、フロリダ大学もこのプロジェクトに参画し、2014 年 6 月の時点で 20 以上のアルゴリズムが実装されている。開発元である Pivotal のデータサイエンティストチームは、主要なデータ分析ツールとしてこの MADlib を採用しており、現在も研究開発が積極的に進められている。なお、MADlib はオープンソースツールとして、ウェブサイトからダウンロードが可能である [7]。以上のような研究開発業

務はデータサイエンティストの必須業務に定められているわけではないが、Pivotal データサイエンスチームでは研究開発に積極的に関わることが奨励されている。

6. Pivotal データサイエンスチーム内の情報共有

最後に、地理的に離れた複数の拠点を持つデータサイエンスチームが取り組んでいる 2 つの情報共有の方法について紹介する。1 つはオンライン上のファイル共有システムである。Pivotal データサイエンスチームは、Pivotal 内でも彼らだけにアクセスが許されている専用ウェブページを持ち、これまでの活動の軌跡を蓄積している。具体的には、これまで関わったデータ分析事例の要約、分析アルゴリズム、分析ツール開発関連資料等がすべての Pivotal データサイエンティストに共有されている。分析プロジェクトやツール開発に携わった者が随時ウェブページに資料をアップロードする形で知識の蓄積が行われている。この蓄積された情報を利用し、新たな顧客案件に必要な知識を学ぶことができるので、データサイエンスチーム全体のスキルアップにつながっている。

もう 1 つ情報共有手段は、電話（テレビ）会議である。電話会議の目的は主に 2 つあり、1 つは Pivotal データサイエンティストが関わっているプロジェクトの進捗状況を共有すること、そして各プロジェクトで直面している課題について相談すること、の 2 点である。電話会議では、各拠点のデータサイエンティストが一堂に会し新しいデータ分析案件の情報交換、新規の分析手法についての意見交換が行われる。時差の関係があるため、毎回 50 名全員が参加することはないのが現状だが、毎回少なくとも 25 名は参加している。関わっているデータ分析関連プロジェクトで直面している課題があったとしても 25 名のデータサイエンティストからアドバイスをもらえる環境があることで、課題を解決するまでの時間がかなり短縮される。地理的に離れていても 1 つのチームとして互いに協力し合っているのを感じられるのがこの電話会議の場である。

7. おわりに

本稿では、グローバルに活躍するデータサイエンティスト集団の例として、弊社 Pivotal の事例を取り上げ、その活動内容を紹介した。グローバル企業のデータサイエンティストは、ビジネス課題に直結するデータ分析プロジェクトに携わり顧客企業への貢献するのはもちろん、サイエンティストというアカデミックな側面も

あわせ持ち、学术界においても業績を残している。彼らの活動がビジネスと学术界をつなぐ架け橋的な役割になる可能性は無視できない。またその研究開発の土壌は日々の情報共有の場で培われており、50名のデータサイエンティスト集団は、他企業の研究開発部門と同じく、日々新たな分析手法の開発に力を入れている。このようなデータサイエンスチームは日本ばかりか、アメリカ国内にも例を見ない。その意味で、今後日本のデータ分析を武器とする企業がグローバル展開をする際に、Pivotal データサイエンスチームの取組みは先行事例として考慮するに値する。本稿で紹介したPivotal データサイエンスチームの取組みが、今後データサイエンティストになることを目指す方々にとって参考になれば幸いである。

参考文献

- [1] <https://dsi.virginia.edu/academics>
- [2] EMC トレーニングの Data Science and Big Data Analytics, http://japan.emc.com/collateral/support-training/training/customer-training-offerings/Data_Science_and_Big_Data_Analytics.pdf
- [3] EMC トレーニングの Data Science and Big Data Analytics for Business Transformation, http://japan.emc.com/collateral/support-training/training/customer-training-offerings/1Day_Data_Science.pdf
- [4] D. Lin, R. Raghu, V. Ramamurthy, J. Yu, R. Radhakrishnan and J. Fernandez, “Unveiling clusters of events for alert and incident management in large-scale enterprise IT,” In *Proceedings of SIG KDD 2014*, in press.
- [5] J. Veness, K. S. Ng, M. Hunter, W. Uther and D. Silver, “A monte-carlo AIXI approximation,” *Journal of Artificial Intelligence Research*, **40**, 95–142, 2011.
- [6] Journal of Artificial Intelligence Research, <https://www.jair.org/bestpaper.html>
- [7] <http://madlib.net>