

データサイエンティストを辞めさせる方法



Ingo Mierswa

データサイエンティストを管理するにあたり、少なくとも正しいことを一つ行っている組織はあるようです。それは、彼らが他の職場を探すのに十分な理由を与えることです。データサイエンティストがそんなに素晴らしい職なら、この転職市場の大量離職をどう説明すればいいのでしょうか。

そこで、もしあなたがこれを読んで、「今、データサイエンティストの離職率が本当に高いと思うのは私だけだろうか」と思っているなら。

教えてあげましょう。いいえ、あなただけではありません。データサイエンティストの離職率はこれまでになく高くなっています。実際に、[365 Data Science Ltd](#) による調査では、平均的なデータサイエンティストは 1.7 年しか仕事を続けられないことがわかりました。

[労働統計局](#) はデータサイエンスの雇用成長率が 30-35% 増加すると予測しており、これは 2020 年から 2023 年の間に 15,000 以上もの雇用機会が生み出されることを意味しているため、仕事に対する不満の蔓延と、データサイエンティストの高い需要とは全く対照的になっています。

私自身データサイエンティストとして、このことはデータサイエンティストが [ユニコーン](#) のようにユニークであった古き良き時代をなつかしく思わせるほどです。それが、今や動物園のように感じられます。たくさんの異なる肩書きや役割によって混雑してしまい、採用するだけでなく、現在いるデータサイエンスの人材を維持することにも激しい競争があります。

では、何が、データサイエンティストがそろって辞める原因なのでしょう。それを探っていきましょう！

もちろん、私はデータサイエンスの古き良き日々を懐かしんでいるわけでは**ありません**。自然言語処理のために Perl でコーディングするような時代に戻りたい人はいるでしょうか？ もちろん、私はそうは思いません



データサイエンティストを辞めさせたいなら行うべき 6 つのこと

今日、データサイエンティストはあらゆる権力を持っています。データサイエンティストに対する需要は供給を大きく上回っており、彼らが現在の組織に不満があれば、辞めることができます。データサイエンティストが去っても問題ないなら、以下の 6 つのことを実行するようにしましょう。

1. データサイエンティストに成長機会がゼロの仕事を与える

よくある誤解に、データサイエンティストは永遠にデータサイエンティストでいることに満足している、とうものがあります。そのために学校に行ったのだから？ データサイエンスを行うために。夢のような生活を送っているに違いない！と。

いいえ。

良いデータサイエンティストは数学者か行動派ですが、彼らは全員根っからの問題解決者です。明確なキャリアパスや従事すべき新しい問題がなければ、彼らの野心の行き場はなくなってしまうでしょう。

データサイエンスは絶えず進化していきます。そこには常に、[機械学習モデル](#)の作成や、ビジネス課題の解決のために、新しいツールやプラットフォーム、メソッドがあります。学習し続け、個人的な成長を促し、組織のためにベストな結果を出し続けるためには、彼らにカンファレンスに参加させたり、さらなるトレーニングを受けさせたり、好みの環境で仕事をさせるなどの機会を与えることが重要です。

またデータサイエンスの役割にはかなり多様性があり、統計学、データリテラシー、コンピュータサイエンス、ストーリーテリング、プレゼンテーション、リーダーシップなどのスキルを必要とします。彼らのキャリアパスは明確なものではなく、むしろ彼ら特有の専門性を基にしたユニークな旅路なのです。

もし現在の会社でデータサイエンティストとして成長する余地がなく、雇用主が彼らを成長させる方法がわからないなら、なぜ彼らは自身の才能を別の場所へ持って行かないのでしょうか。

2. データサイエンティストの仕事を理解する努力をしない

次のようなことを想像してみてください。今日が世界的な小売企業でデータサイエンティストとして勤務する最初の日です。あなたは簡単な自己紹介をし、数人の同僚と会い、それから最初の業務を与えられます。離反の危険信号を見極める、顧客の離反モデルを作成し、最終的には ROI を改善する、というものです。

これは小さな業務ではなく、この時点で全部で 3 時間しか仕事に従事していません。解こうとしているユースケースの背景をまったく理解しないで、どうやって価値のある機械学習モデルを作成できるのでしょうか。

このようなとき、重役はデータサイエンティストが日々行う業務を理解しない(したがって、信用もしない)のに、なぜデータサイエンティストはビジネスのユースケースの理解を期待されるのか。

このように、データサイエンスを理解する努力がないと、「ビジネスの人」と「データの人」間のスキルギャップをますます広げるという事態になっています。

単純に重役たちがその背後にあるプロセスを理解できないので、より高度なメソッドは初めから考慮されず、複雑なディープラーニングの結論が評価されないと、データサイエンティストたちは自身の時間を無駄にしたように感じ、不満を抱きます。そのようなわけで、ステークホルダーやビジネスリーダーの[アップスキリング](#)が今までよりさらに重要になります。

3. データサイエンティストに退屈な作業をさせる

データサイエンティストに「なぜデータサイエンティストになったのですか？」と尋ねても、「一日中データの前処理を行いたかったからです！」と答える人はおそらくいないでしょう。

先人のデータサイエンティストの最近の [Medium](#) の記事で、この点が証明されました。この記事では、時間の 90% を費やしたかったモデルの作成が、実際には 10% の時間しか費やせなかったと話しています。また彼はデータサイエンスについて、正直に言うと、彼が思っていたよりも退屈で、やりがいや充実感は期待していたものの半分もないと投稿していました。

データサイエンティストに一日中、Excel 内でセルや列を並び替えたり、古い手法を使った単純な分析をさせるために彼らを雇ったわけではありません。では、なぜ彼らはそれほど多くの時間を、回避可能で、つまらないデータの前処理や整理に費やしているのでしょうか。

実際には、データサイエンティストの時間の多くが乱雑なデータの整理や、論争、集計、整形、SQL クエリの実行、実行したコードの待ち時間に費やされています。これにより、たくさんのデータサイエンティストが輝かしいユニコーンというよりも、データの管理人さんのように感じています。

誤解のないようにですが、データの前処理は、データサイエンスにおいて非常に重要なパートです。ただし、データの前処理のみにチームをやらせないようにしてください…

4. データサイエンスに特化したツールの使用をすべて拒否する

データサイエンティストは、試行錯誤を繰り返す人たちです。このような人たちは、市場で最も良い、最新のツールを与えられたときに最も良い働きをします。

残念なことに、多くの企業が良いデータサイエンスツールに投資をするわけではありません。中には、全く投資しない企業もあります。専用のデータサイエンスプラットフォームを持たないと、既存のファイアウォールに立ち向かい、実際に[モデルを運用していく](#)のは非常に難しくなるでしょう。

たとえ組織がプログラムを持っていたとしても、十分なトレーニング(RapidMiner [Academy](#) の認定のようなもの)がなければ、プラットフォーム内で自身のスキルを適用し、伸ばす方法がないのです。また、補助を行うデータエンジニアリングツールに投資せず、データサイエンティストに非構造で一貫性のないデータの整理をさせる組織もあります。(データサイエンティストがデータの退屈な作業にどのように感じるかは、上記をもう一度見てください。)

データサイエンティストに業務上必要最低限のツールしか与えないと、従業員たちは経営者が自分たちのベストな成果を出せるような技術的な環境に投資する意思がないのだとわかります。

5. かわりに、手作業や、非効率な手法、ソフトウェアしか使用させない

また同じくらい悪いこととして、全くツールを気にしないことがあります。彼らが使いたいものを使わせましょう。なぜなら、多数の異なる監査様式に対応できるような、合理的で、受容できる、堅牢な方法はないため、禁止することで、彼らのモデルのほとんどを行き詰らせ、けっして運用に回せるようにはならないでしょう。これは「**彼らは何の成果も上げていない**」ということを意味します。

ツール市場は、ただデータの前処理やモデリングに使用されるツールにとどまりません。データサイエンティストに、彼らの使用する言語を用いるソフトウェアを与えることは、生産性を最大限に上げるキーです。しかし、いまだに PowerPoint のスライドを編集して自身の成果を発表することに時間を費やしているデータサイエンティストの数に驚くでしょう。

データサイエンスの結果を表示、提示する合理的な方法がないため、ビジネスとデータ専門家との間にさらに不必要な障害が生まれています。

多くのデータサイエンティストが嘆いている別の問題に、組織にバージョン管理が不足しているというのがあります。ビジネスの場では、いまだにコードをメールや Google Drive、SharePoint で共有しています。コ

ードへの安全性が低だけでなく、仲間の開発者によって書き換えられる可能性があります。つまり、せっかく苦労した作業が知らず知らずのうちに変更されていたり、失われる可能性があります。

作業内容に安全性が低く、本来より 2 倍の時間がかかるなら、データサイエンティストは 100%の満足はしていないでしょう。

6. 実際の価値より低い報酬を支払う

いくらあなたが自身の仕事や同僚、企業を好きでも、給料が低いなら辞めることに躊躇はありませんよね？

残念ながら、データサイエンティストは低い給与に悩まされることが多いです。ML library PyCaret の創業者が [LinkedIn の投稿](#)で、ある例に注目を集めました。「エントリーレベル」のデータエンジニアの求人は、6 年以上の経験を求めているのに、年俵はたったの 62,000 ドルでした。

これは、インパクトのあるツールを設計・開発し、堅牢な技術フレームワークを組織に実装し、ソリューションをステークホルダーに提示することを求められる人物の給与が低すぎるだけでなく、将来の雇用主がこれらの引く手あまたのスキルの価値を全く理解していないことも示しています。あなたは、自分の貢献ぐあいを理解していない人の下で働きたいと思いますか？

データサイエンスの求職者の中には、求人検索で低い給与でフィルターをかける人もいるため、あなたの求人は見られることもないでしょう。または、求人は見ても、求められていることと提供されるものの違いに呆れ、自身の時間(と専門性)を活かすのに別の場所を選ぶ人もいるでしょう。

まとめ

最近の [KDnuggets](#) の記事で、独学でデータサイエンティストになった Natassha Selvaraj は、彼女が学習に費やした年月分のスキルと仕事内容が見合っていないため、インターンをたった 1 週間で辞めたと投稿しています。

このメッセージは明確です。雇用主がデータサイエンティストに投資や、技術的に進んだ仕事環境を提供せず、価値の高いプロジェクトに貢献させなければ、彼らは去っていくでしょう。自分の組織のユニコーンたちが、今よりも良い環境に旅立っていくのを見たい人はいないでしょう。

しかし、リーダーシップをもってコミュニケーション方法を改善することは、データサイエンティストとして私たちにも責任があります。もしあなたが、自身のモデルの価値を明確にすることや、プロジェクト内で信頼を築くのに苦労しているなら、私たちはその手助けをするためのガイドを以下にまとめました。

<https://rapidminer.com/downloads/present-machine-learning-models/>