

# 転職支援におけるデータ解析の活用

～ Visual R Platform による統計分析環境強化 ～

2014年11月21日

株式会社インテリジェンス キャリアBITA部 山之下 拓仁

# 目次

---

## **Section**

**Section 1**            転職支援サービスについて

**Section 2**            データ活用における課題と解決策

**Section 3**            データ解析による課題解決

**Section 1**      **転職支援サービスについて**

---

**Section 2**      **データ活用における課題と解決策**

**Section 3**      **データ解析による課題解決**

## 会社説明①～会社概要～

■社名	株式会社インテリジェンス (英文: Intelligence, Ltd.)
■本社所在地	〒100-6328 東京都千代田区丸の内2-4-1 丸の内ビルディング27F・28F
■資本金	98億1,255万円
■設立年月日	1989年6月15日
■従業員数	単体 3514名 連結 4250名 ※2012年3月31日現在
■サービス内容	DODA人材紹介サービス DODA求人情報サービス 就職支援・採用支援サービス 人材派遣サービス ITソリューションサービス (エンジニア派遣サービス) ITソリューションサービス (アウトソーシングサービス) 「an」サービス 「salida」サービス 組織・人事コンサルティングサービス



**DODA**

**an**

salida サリダ

**CRE DENCE**



# 会社説明②～主要サービス～

## DODA

業種・職種ごとの専任のキャリアコンサルタントがマッチングを行う人材紹介と、約200万人の会員を有する求人サイト「DODA」で質・量ともに幅広いアプローチが可能です

## an

認知度No.1の「an」。女性ターゲットの採用メディア「salida」。多様化する仕事探しニーズに応え、WEBと情報誌で展開

## インテリジェンスの 人材派遣

人材派遣から紹介予定派遣まで充実の福祉校正とフォロー体制であなたに合った働き方をサポート

## 会社説明③ DODA転職支援サービス



転職支援は、転職希望者と求人企業を「結びつける」サービスです

**Section 1**      転職支援サービスについて

**Section 2**      データ活用における課題と解決策

---

**Section 3**      データ解析による課題解決

## データ活用における課題

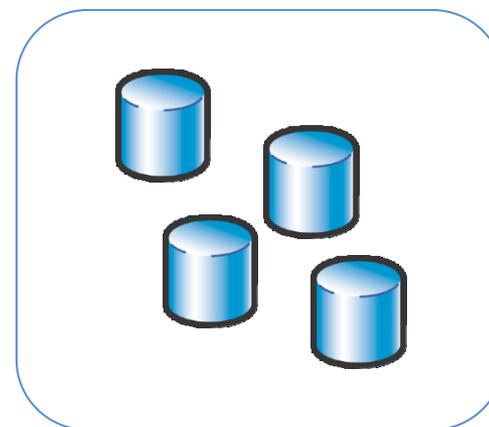
### 課題①

分析ソフトの性能が悪い。  
高度な分析が出来ない。



### 課題②

データが散在。  
活用しきれていない。

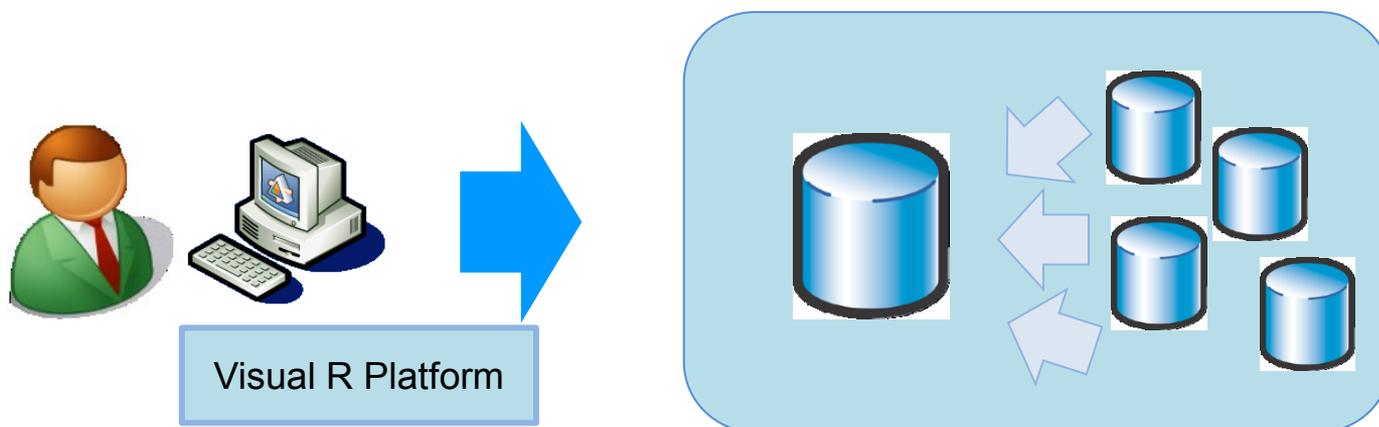


転職希望者の情報と、求人の情報を活用できてない  
→ 転職支援サービスの「質」の向上の余地あり

## データ活用の課題の解決策

### 課題①②解決策

分析環境を整備。散在しているデータを集約。  
高度な統計分析の実現のため、「VRP」を導入。



転職希望者の情報と、求人の情報を有効活用  
→ より精度の高い転職支援サービスを提供

## Visual R Platformの選定理由

---

### 《データ活用に向けた要求事項》

1. 多種多様な最新の分析手法が利用出来る事
2. 近年活用が進んでいるR言語をベースとしている事
3. プログラミングに不慣れなユーザーでもGUIで操作出来る事
4. 高レベルのサポートが期待できる事
5. 今回の要件にライセンス形態が適合している事
6. 比較的低コストで導入できる事



要求事項を満たすソリューションとして「Visual R Platform」を選定

**Section 1**      転職支援サービスについて

**Section 2**      データ活用における課題と解決策

**Section 3**      データ解析による課題解決

---

## 事例紹介前提

---

9月末現在、本格的な利用開始前の為、ユーザ受入テストを兼ねた簡単な分析のご紹介となる点はご了承下さい。

また、公開出来ない箇所はデータの加工を行っており、一部分かり辛い箇所がある点についてもご了承下さい。

この資料について、短期間で用意した為、後日の加筆修正等により当日は同一の物でない可能性がある点についてもご了承下さい。

## 分析事例①: 書類通過の可能性の高い候補者特性の判別と予測

---

### <課題>

個人の方(候補者)が、求人案件を確認して、求人に応募します。  
企業の採用担当者の方々は、履歴書等の提出書類を元に、  
採用選考を進めるかどうか(書類通過)を判断します。

そこで、書類通過の回数を、企業に応募した回数で割った値を書類通過率とし、  
候補者の情報からどの程度予測できるか判別してみます。

※短期間の為、簡易的に使える情報のみで予測しております。

## 分析事例①:書類通過の可能性の高い候補者特性の判別と予測

### <VRPの設定>

今回は、VRPの受入テストとして、基本機能で決定木とニューラルネットワークを、また、R言語でランダムフォレストを利用して通過率を予測をします。まず元データの80%を教師データとして学習させ、残り20%を検証データとして学習したモデルで予測し、正解率を確認します。

The screenshot displays the Visual Analytics Platform interface. The main workspace shows a workflow diagram with the following components and connections:

- 候補者データ** (Candidate Data) feeds into **ランダムフォレスト** (Random Forest) and **ファイル分割** (File Split).
- ランダムフォレスト** feeds into **予測** (Prediction).
- ファイル分割** feeds into **決定木** (Decision Tree) and **ニューラルネットワーク** (Neural Network).
- 決定木** and **ニューラルネットワーク** both feed into **予測** (Prediction).

The **VRP ファイル分割** (VRP File Split) dialog box is open, showing the following settings:

- 分割比** (Split Ratio): 学習用 (Learning) is 8, 検証用 (Validation) is 2.
- 乱数の初期値** (Random Seed):  自動 (Automatic),  手動 (Manual) with a value of 0.

The Message pane at the bottom shows the following log entries:

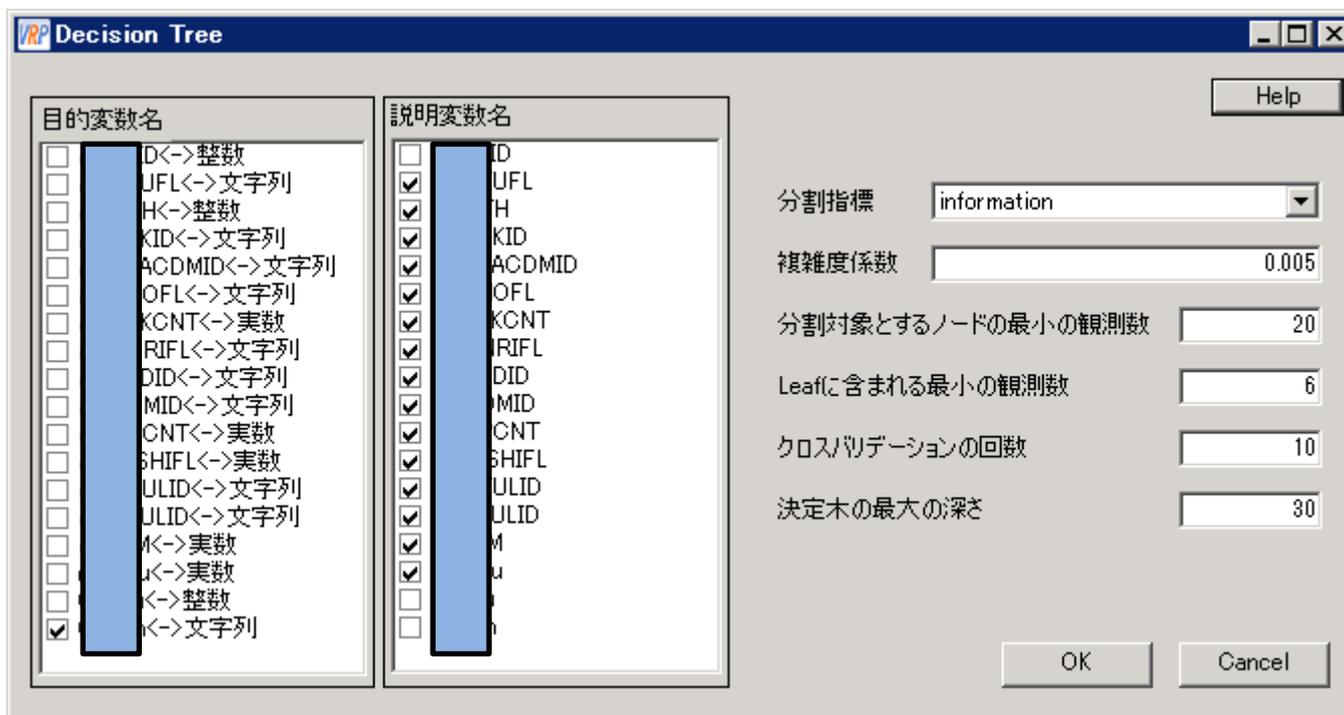
```

データ読み込み開始 ...
データの読み込み完了
RデータのVRP出力開始 ...
RデータのVRP出力完了
  
```

## 分析事例①: 書類通過の可能性の高い候補者特性の判別と予測

### <VRPでの予測①: 決定木>

書類通過率を上位50%をA、下位50%をBとして決定木で予測してみます。



## 分析事例①: 書類通過の可能性の高い候補者特性の判別と予測

### <VRPでの予測①: 決定木>

正解率が64%程度でしたが、要因が人が見て分かりやすく、処理が速いので、概要を把握するには良い感じでした。

Decision Tree --- データ&グラフィックビュー

resultModel for R    resultCPTable for R    **resultDT for R**

行名	Node	Path	Split	N	Loss	Value	Is.Leaf
1	1	1	root	63216	27994	B (0.442830928878765 0.557169071121235)	FALSE
2	2	1-2	< 0.5	23154	9969	A (0.569448043534594 0.430551956465406)	FALSE
3	3	4 1-2-4	< 1,3,7,9,10,99	14588	5070	A (0.652454071839868 0.347545928160132)	TRUE
4	4	5 1-2-5	< 2,4,5,6,8,11	8566	3667	B (0.428087788932991 0.571912211067009)	TRUE
5	5	3 1-3	< 0.5	40062	14809	B (0.369652039339025 0.630347960660976)	FALSE
6	6	6 1-3-6	< 1,3,6,7,9,10	21937	9525	B (0.434197930437161 0.565802069562839)	FALSE
7	7	12 1-3-6-12	< 0.5	7729	3464	A (0.551817828955880 0.448182171044120)	TRUE
8	8	13 1-3-6-13	>= 0.5	14208	5260	B (0.370213963963964 0.629786036036036)	TRUE
9	9	7 1-3-7	< 2,4,5,8,11,99	18125	5284	B (0.291531034482759 0.708468965517241)	TRUE

予測 --- データ&グラフィックビュー

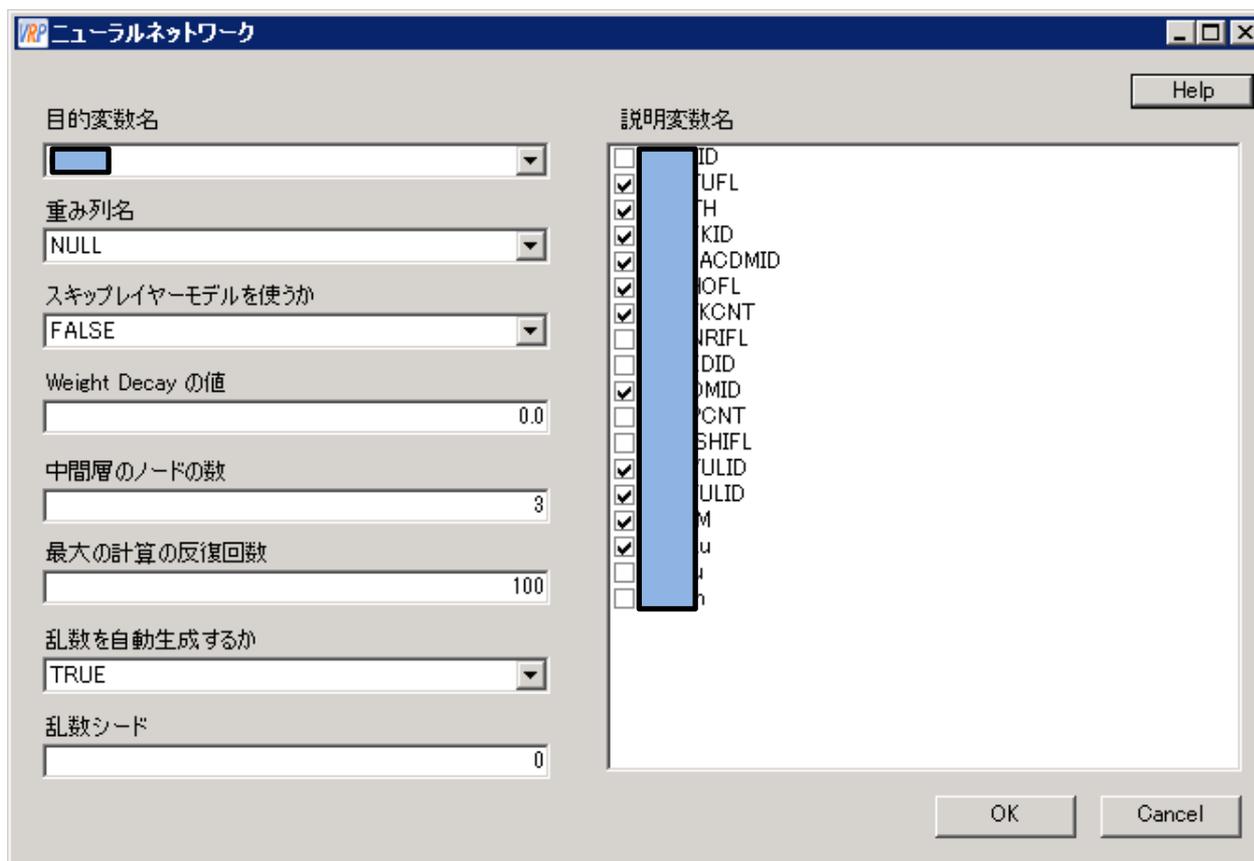
result for R    extraValue for R    summary for R    **precision for R**    recall for R

行名	Prediction.A	Prediction.B	Prediction.Sum
1 A	62.426	34.502	44.584
2 B	37.574	65.498	55.416
3 Sum	100.000	100.000	100.000

## 分析事例①:書類通過の可能性の高い候補者特性の判別と予測

### <VRPでの予測②:ニューラルネットワーク>

書類通過率を3段階に、A、B、Cに分けてニューラルネットワークで予測してみます。



## 分析事例①:書類通過の可能性の高い候補者特性の判別と予測

### <VRPでの予測②:ニューラルネットワーク>

今回は時間の都合上、変数を減らして、チューニングもしていないので正解率が66%程度になりましたが、さらに改善できそうです。

result for R		extraValue for R	summary for R	precision for R	recall for R
行名	Prediction.A	Prediction.B	Prediction.C	Prediction.Sum	
1 A	92.982	15.575	33.333	15.895	
2 B	1.754	61.015	22.222	60.712	
3 C	5.263	23.410	44.444	23.393	
4 Sum	100.000	100.000	100.000	100.000	

resultModel for R		resultSummary for R
行名	coef.result.model.	
1 b->h1	101.470	
2 i1->h1	1.285	
3 i2->h1	-28.811	
4 i3->h1	-1.576	
5 i4->h1	-0.235	
6 i5->h1	0.190	
7 i6->h1	1.492	
8 i7->h1	0.608	
9 i8->h1	0.528	
10 i9->h1	0.535	
11 i10->h1	24.500	
12 i11->h1	11.955	
13 i12->h1	-2.787	
14 i13->h1	15.989	
15 i14->h1	-58.731	
16 i15->h1	-41.907	
17 i16->h1	-64.255	
18 i17->h1	0.301	
19 i18->h1	-0.490	
20 i19->h1	-69.798	

## 分析事例②:書類通過の可能性の高い候補者と求人案件のマッチング

---

### <課題>

インテリジェンスでは「インフラとしての人材サービスを提供し、はたらくを楽しむ社会を実現する」というビジョンを持っており、DODAのサービスを通じ、弊社コンサルタントや、メールでのご提案により、候補者の方々に対し、よりはたらくを楽しめる求人案件の提案を行っております。

そのうち、DODAによって採用が「決まる」提案をしていく事をIT技術で支援する為、機械学習を適用する事でより効率よく提案していけるか検討してみます。

利用イメージとして、機械学習により「決まりやすい」案件を絞り込み、その中からコンサルタント等が求人を提案する事を想定します。

## 分析事例②: 書類通過の可能性の高い候補者と求人案件のマッチング

### <KPIの設定>

教師データとして8月末以前の約一年分の候補者と求人案件の通過実績データ、  
検証データとして、9月以降の約一か月分のデータを用意し、  
Random Forest による予測モデルで検証データの予測値を出しました。

実績値の書類通過の○×と予測値の○×を比較して、  
正解率(①+②÷(①~④の計))・・・ $p_1$ が高い程よいですが、  
実用面においては、書類通過するかもしれないマッチングをカバー出来る様、  
書類通過○の時の再現率(①÷(①+②))・・・ $p_2$ も重視します。

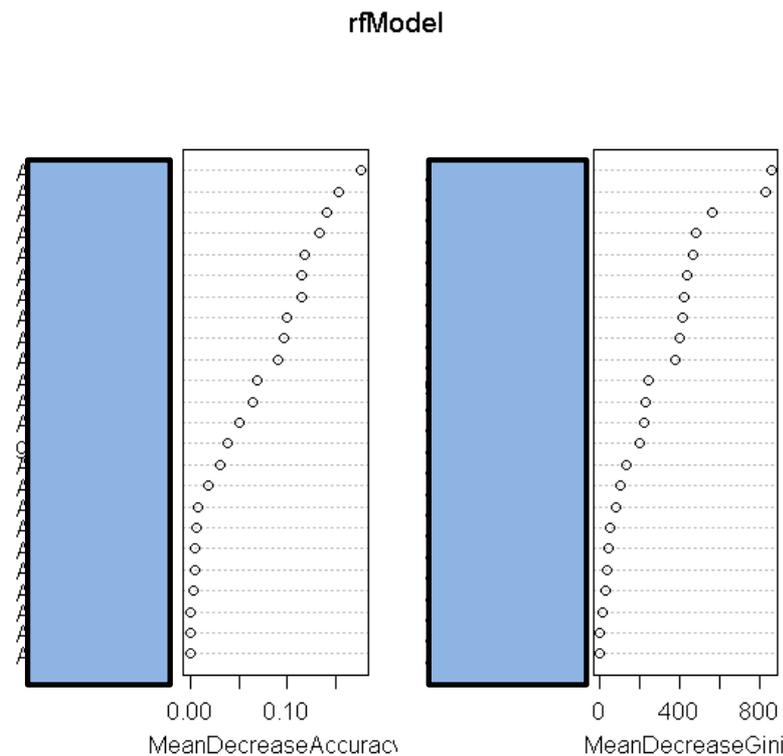
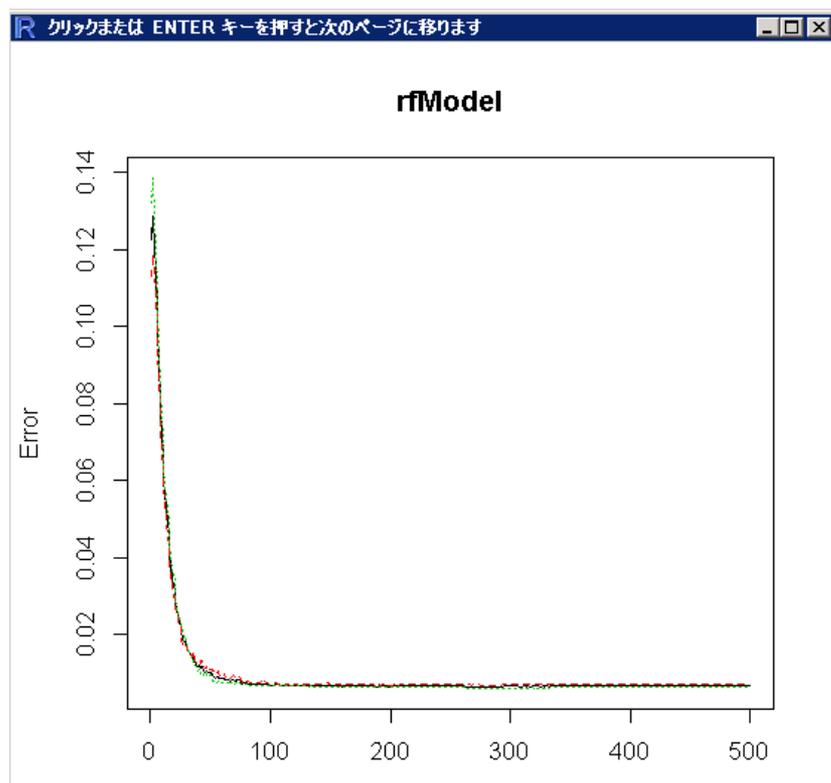
### 予測値

		予測値	
		書類通過○	書類通過×
実績値	書類通過○	①	②
	書類通過×	③	④

## 分析事例②: 書類通過の可能性の高い候補者と求人案件のマッチング

<Rでの予測: ランダムフォレスト>

今回はRで実行しましたが、VRPでもRを実行出来るので同じ結果が得られます。



## 分析事例②: 書類通過の可能性の高い候補者と求人案件のマッチング

<Rでの予測: ランダムフォレスト>

特定グループに絞り、ある程度チューニングした結果、元データに対しては正解率99%、検証データに対しては正解率82%再現率89%となりました。

Call:

```
randomForest(formula = formula.ob, data = input.le, ntree = 500, mtry = 5, importance = TRUE, proximity = TRUE)
```

Type of random forest: classification

Number of trees: 500

No. of variables tried at each split: 5

OOB estimate of error rate: 0.84%

Confusion matrix:

0 1 class.error

0 3699 56 0.014913449

1 30 6439 0.004637502

<予測結果>

0 1

0 205 113

1 90 746

<正解率・再現率>

[1] 0.8240901 0.8923445