

MSIISM Conference 2024

社員の声を経営に活かす。働きがい向上につなげる施策と分析。

# ベイジアンネットワークとテキスト分析による 仮説検証とその深化 ~人事データ分析を例に~

2024年10月11日

株式会社NTTデータ数理システム

岩本 圭介

# データ分析とは「モデル化」である



どんな傾向があるんだ？

何が起きているんだ？

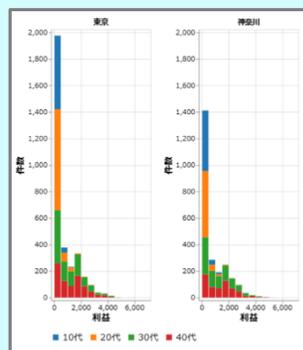
データを通して  
知りたい



**モデル化** はデータに問いかける、データに語ってもらうための手段

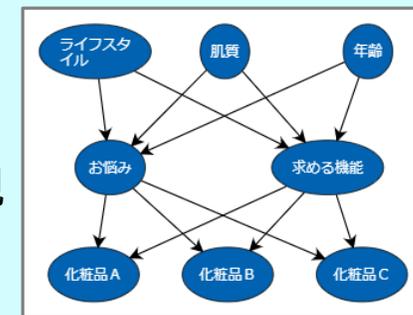
## 集約する

- 可視化による把握
- 類似データをまとめ上げて観察
- 平均値などの統計量算出も、データの代表的な側面をあぶり出す手段



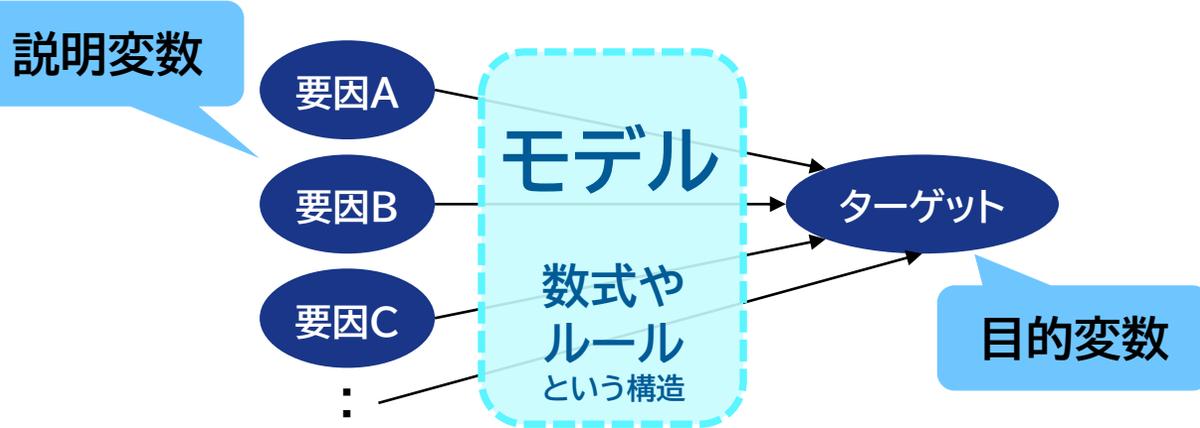
## 構造化する

- 要因とターゲットとの関係を数式・ルールによって近似表現
- 要因間にどのような依存関係の構造があるか探索



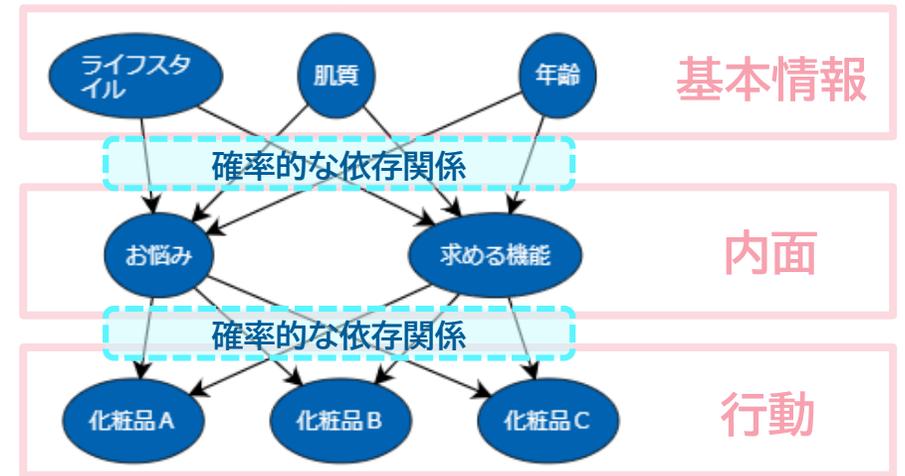
# モデル化により、データに「構造」を見い出す

- 要因とターゲットとの関係を数式やルールによって表現する



- 事象の多段階の連鎖構造を「丸ごと」モデルとして表現する

## ベイジアンネットワーク



いかに緻密な良い近似ができるか？  
という点と  
モデルがいかに人に理解し易いか？  
という点のトレードオフやバランス

いかに実世界の事象の  
発生機序を反映できているか？  
ということを重視しつつ  
それが良い近似であることを目指す

# 「仮説設定」がデータ分析に切り口を与える

※(株)NTTデータ数理システムにて構築した架空の事例をもとにご紹介します

当社は業績も好調に推移しているのですが、  
離職者が増えていることが悩みです。  
社員の行動やアンケートのデータを採取しています。  
データから離職につながる要因を  
解明できないものではないでしょうか？



多くのデータ項目がありますね。  
幸い、対象データは 5種類 に区分できています。  
情報の整理のため、それらの  
区分レベルでの関係について仮説を立ててみましょう。

# 情報を整理し、「依存関係の大枠」について仮説を立てる

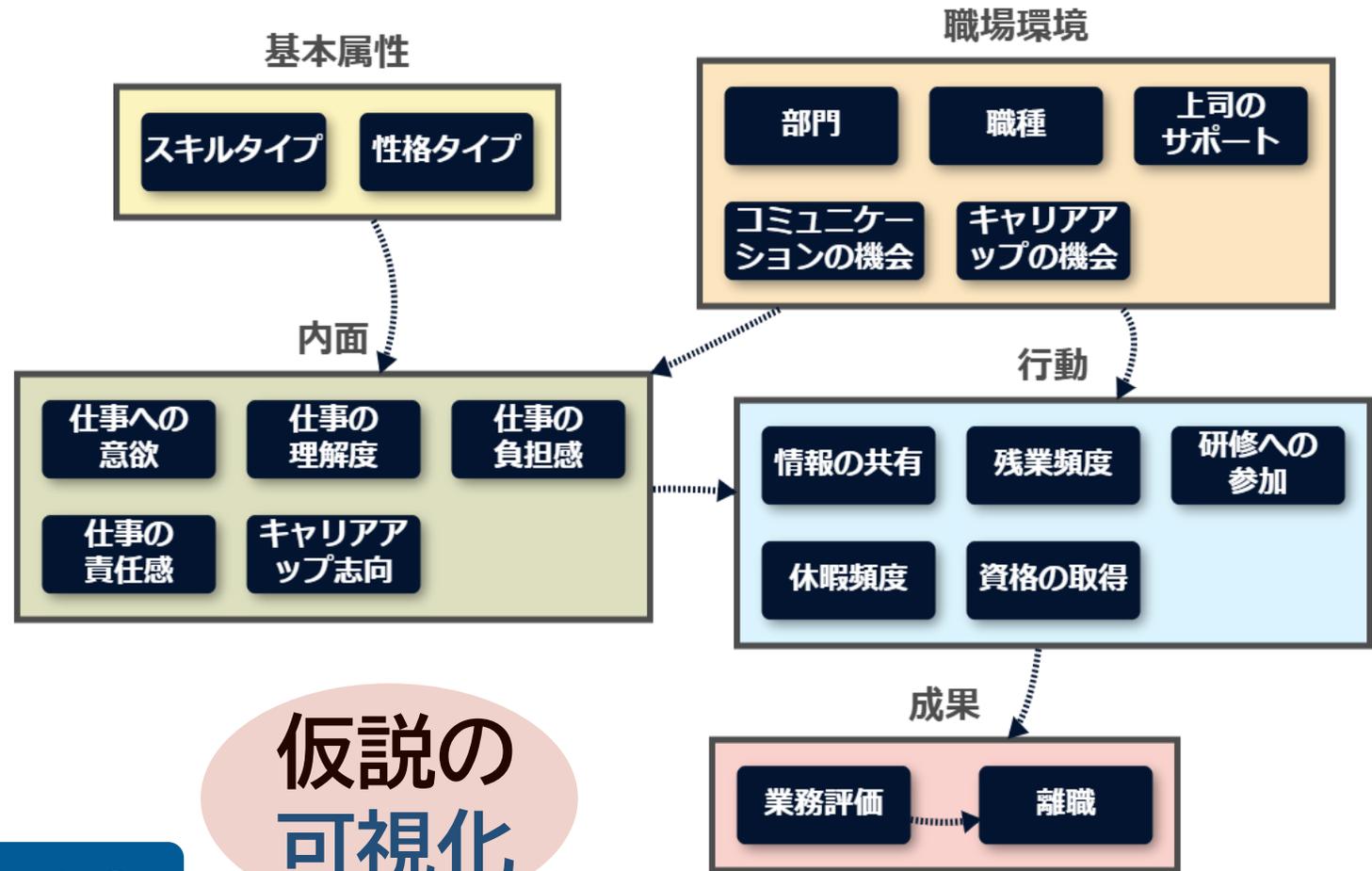


こういった仮説が立てられるのではないのでしょうか？

- **基本属性**は入社前から備わっており、他の要素から影響を受けない
- 配属の際に基本属性は特に考慮されていないとのことなので、**職場環境**は基本属性とは関係しない
- **内面**は**行動**は職場環境や基本属性から影響を受け変化する
- さらに**内面**は**行動**に影響を及ぼす
- そこで決定する行動から、最終的な**成果**が決まる



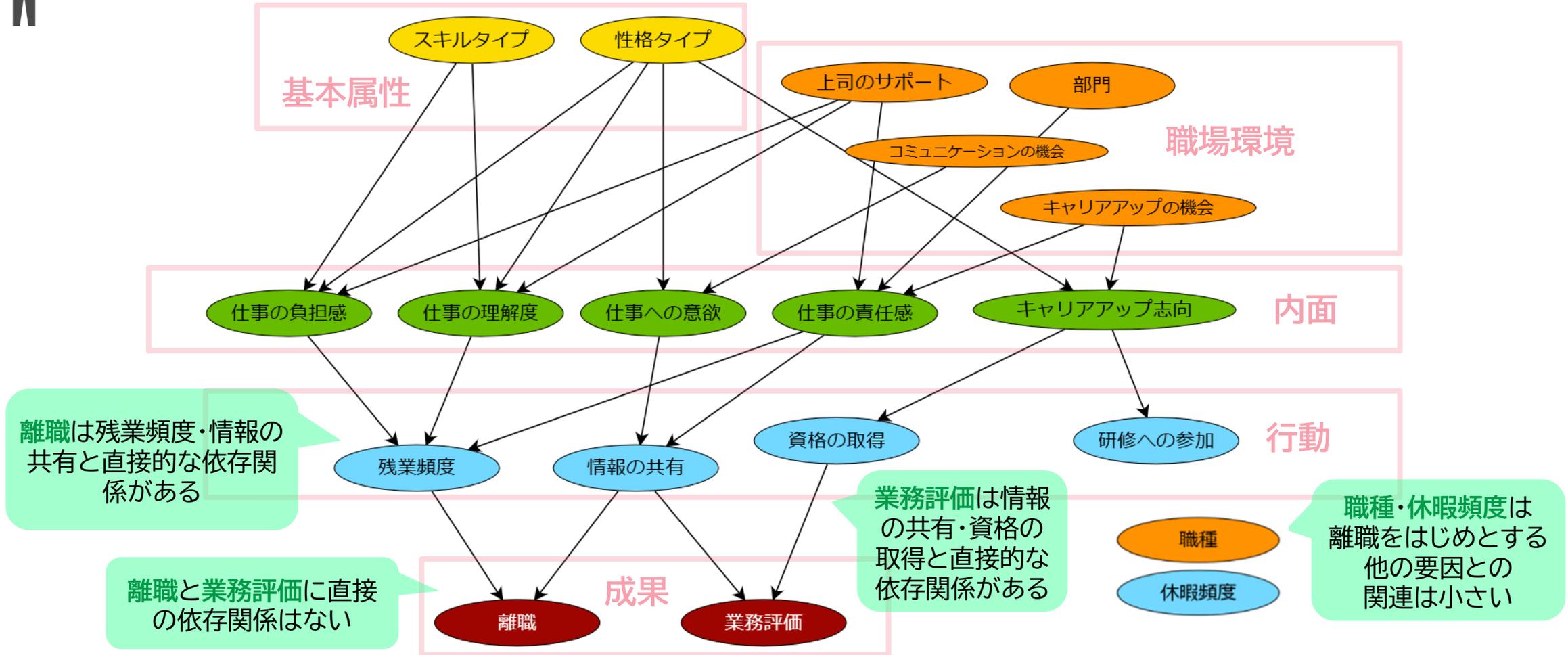
……と考えてみてはどうでしょう？



# 大枠をもとに「構造学習」で依存関係のネットワークを構築する



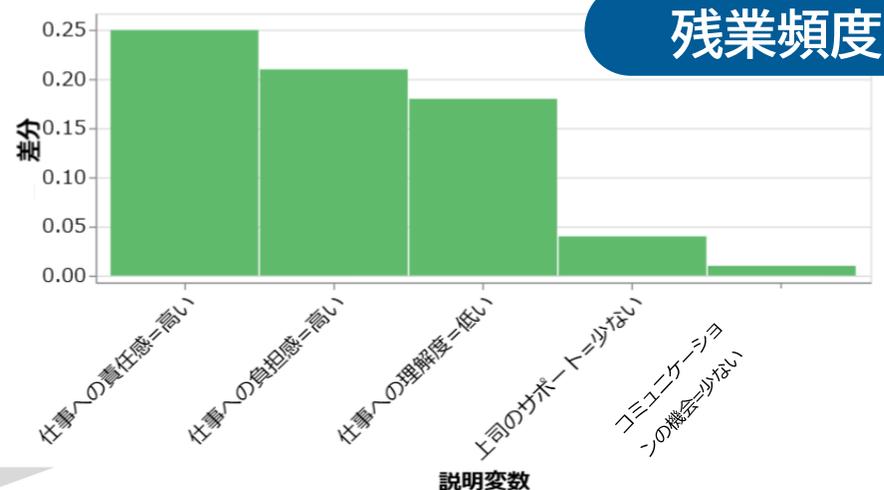
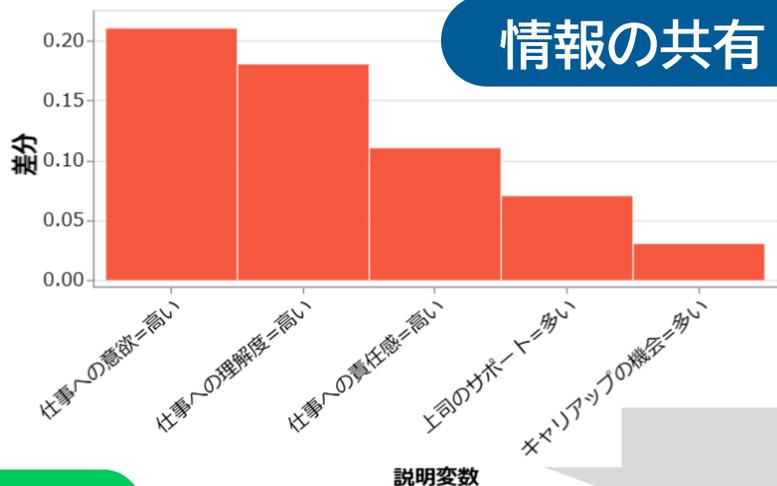
大枠のみを仮定して学習を行った結果、  
次のようなベイジアンネットワークが得られました。



# 「確率推論」で影響を定量的に評価する



さらに、離職と依存関係にある **情報の共有・残業頻度** に影響する因子とその影響度合いを評価しました。



## さらなる仮説



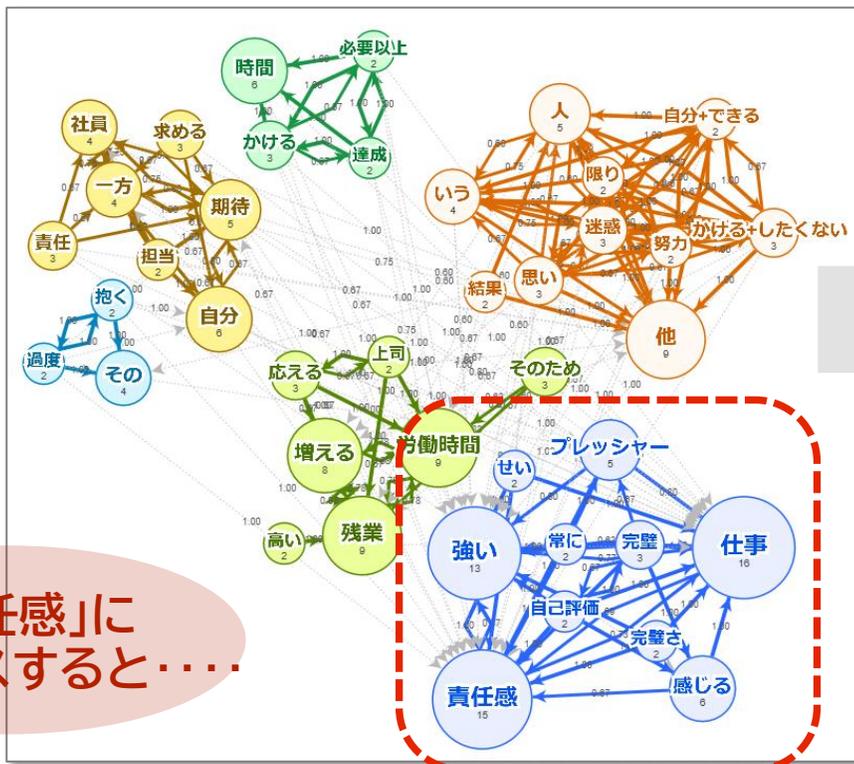
- 責任感が高いと情報の共有など良い働きにもつながるが、残業の要因にもなりうる
- 負担感が高い点や仕事への理解度が低い点を解消することで残業を回避できるのではないか
- 負担感、理解度の両方へ影響を及ぼす上司からのサポートにより改善が見込めるのではないか

データドリブンな  
仮説の  
掘り下げ・深化

# 「テキスト分析」で根拠の裏付けとなる背景事情を明らかにする



続いて、アンケートの自由回答から背景事情の読み解きを試みました。



「責任感」にフォーカスすると……

Alkano「ネットワークからの話題抽出」機能

原文参照	
Row...	原文
9	責任感が強いため、仕事への取り組みに真剣さを求める一方で、その責任感が時にプレッシャーとなってしまう、自分に対して過度な期待を抱いてしまいます。
10	責任感が強いせいで、仕事の成果に対して常に責任を感じるため、完璧な仕事をするというプレッシャーが付きまといまます。
13	責任感が強く、仕事に対して自己評価が高いため、自信が持てないと残業をしてまで仕事を完璧に仕上げようとするプレッシャーを感じます。

- 責任感 はプレッシャーとなってしまうこともある。ただその要因は他者からによるものではなく自己評価に由来していることも多い。

仮説構築



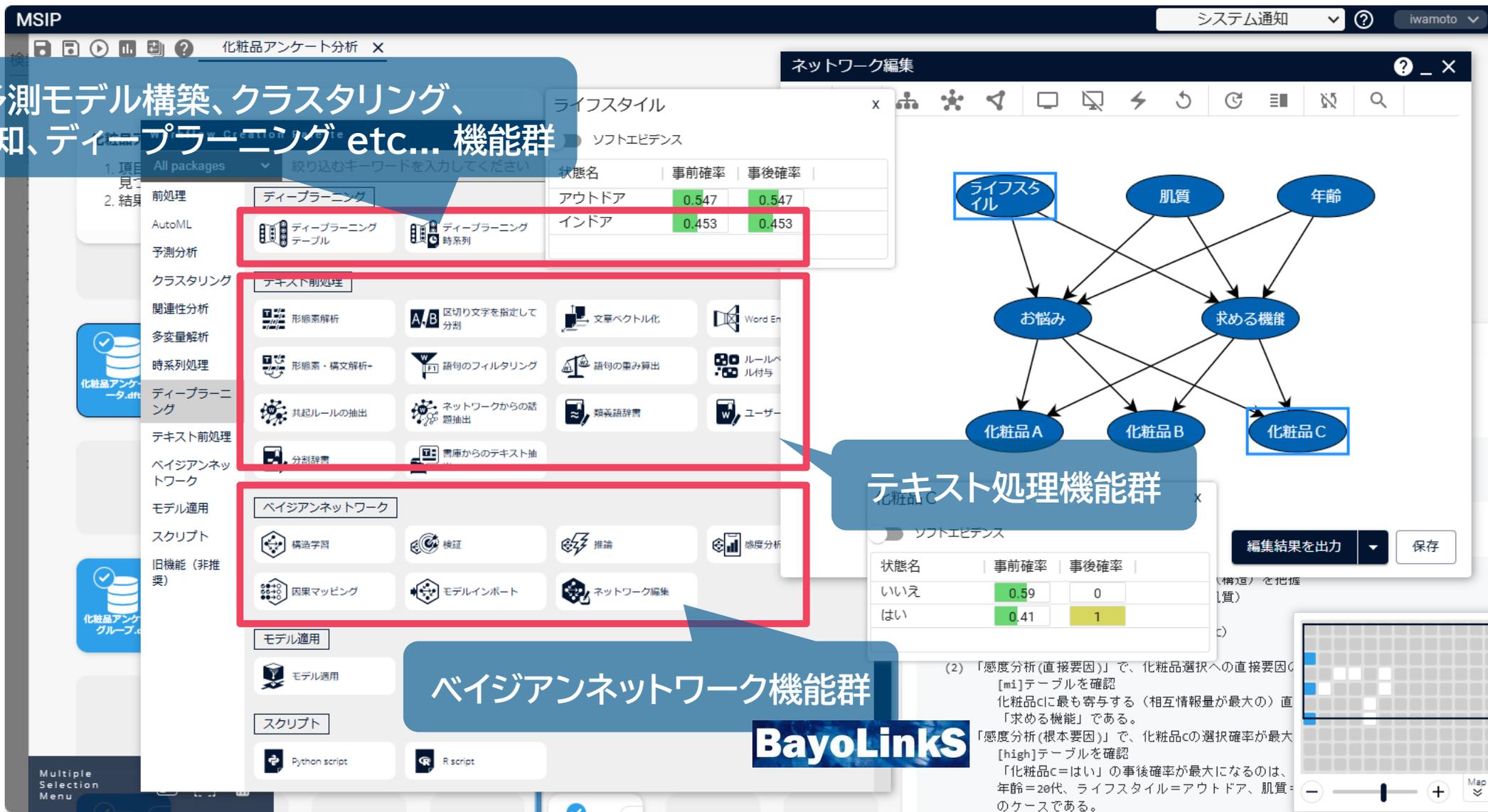
仮説の深化



裏付け検証

# データ分析プラットフォーム Alkano

予測モデル構築、クラスタリング、異常検知、ディープラーニング etc... 機能群



The screenshot displays the Alkano platform interface with several functional groups highlighted in red boxes and labeled with callouts:

- ディープラーニング機能群** (Deep Learning Functional Group): Includes options for 'ディープラーニング テーブル' and 'ディープラーニング 時系列'.
- テキスト処理機能群** (Text Processing Functional Group): Includes various text analysis tools such as '形態素解析', '区切り文字を指定して分割', '文章ベクトル化', 'Word Embedding', '形素素・構文解析+', '語句のフィルタリング', '語句の重み算出', 'ルールレベル付与', '共起ルールの抽出', 'ネットワークからの話題抽出', '類義語辞書', 'ユーザー', and '分割辞書'.
- ベイジアンネットワーク機能群** (Bayesian Network Functional Group): Includes 'ベイジアンネットワーク' and 'ネットワーク編集'.

The interface also shows a 'ネットワーク編集' (Network Editing) window with a directed graph and a 'ソフトエビデンス' (Soft Evidence) table.

**ネットワーク編集**

Graph nodes: ライフスタイル, 肌質, 年齢, お悩み, 求める機能, 化粧品A, 化粧品B, 化粧品C.

**ソフトエビデンス**

状態名	事前確率	事後確率
アウトドア	0.547	0.547
インドア	0.453	0.453

**ソフトエビデンス**

状態名	事前確率	事後確率
いいえ	0.59	0
はい	0.41	1

**BayoLinks**

(2) 「感度分析(直接要因)」で、化粧品選択への直接要因の [mi] テーブルを確認  
化粧品cに最も寄与する(相互情報量が最大の)直接「求める機能」である。  
「感度分析(根本要因)」で、化粧品cの選択確率が最大 [high] テーブルを確認  
「化粧品c=はい」の事後確率が最大になるのは、年齢=20代、ライフスタイル=アウトドア、肌質=のケースである。

# 情報整理・仮説立案～依存関係解明・定量評価 の過程をサポートする機能

因果マッピング

データインポートパネル

列名	型名	グループ名
年齢	category	基本情報
休日の過...	category	基本情報
肌質	category	基本情報
お悩み	category	内面
求める機能	category	内面
化粧品A...	category	行動
化粧品B...	category	行動
化粧品C...	category	行動

基本情報

年齢 休日の過ごし方 肌質

内面

お悩み 求める機能

行動

化粧品Aを購入する? 化粧品Bを購入する? 化粧品Cを購入する?

図形 文字

形状

幅 120 px

高さ 50 px

塗りつぶし

色

透過性 100 %

線

色

透過性 100 %

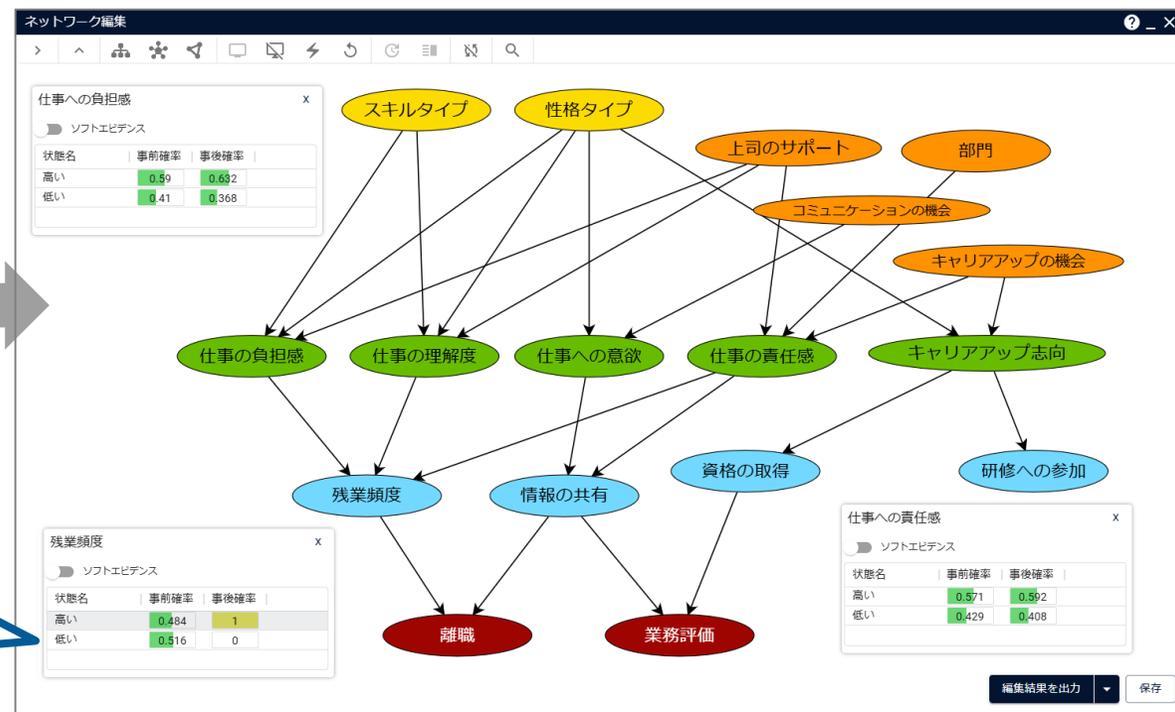
幅 2 pt

実行 保存

情報整理と仮説立案を  
グラフィカルにサポートする  
因果マッピング機能

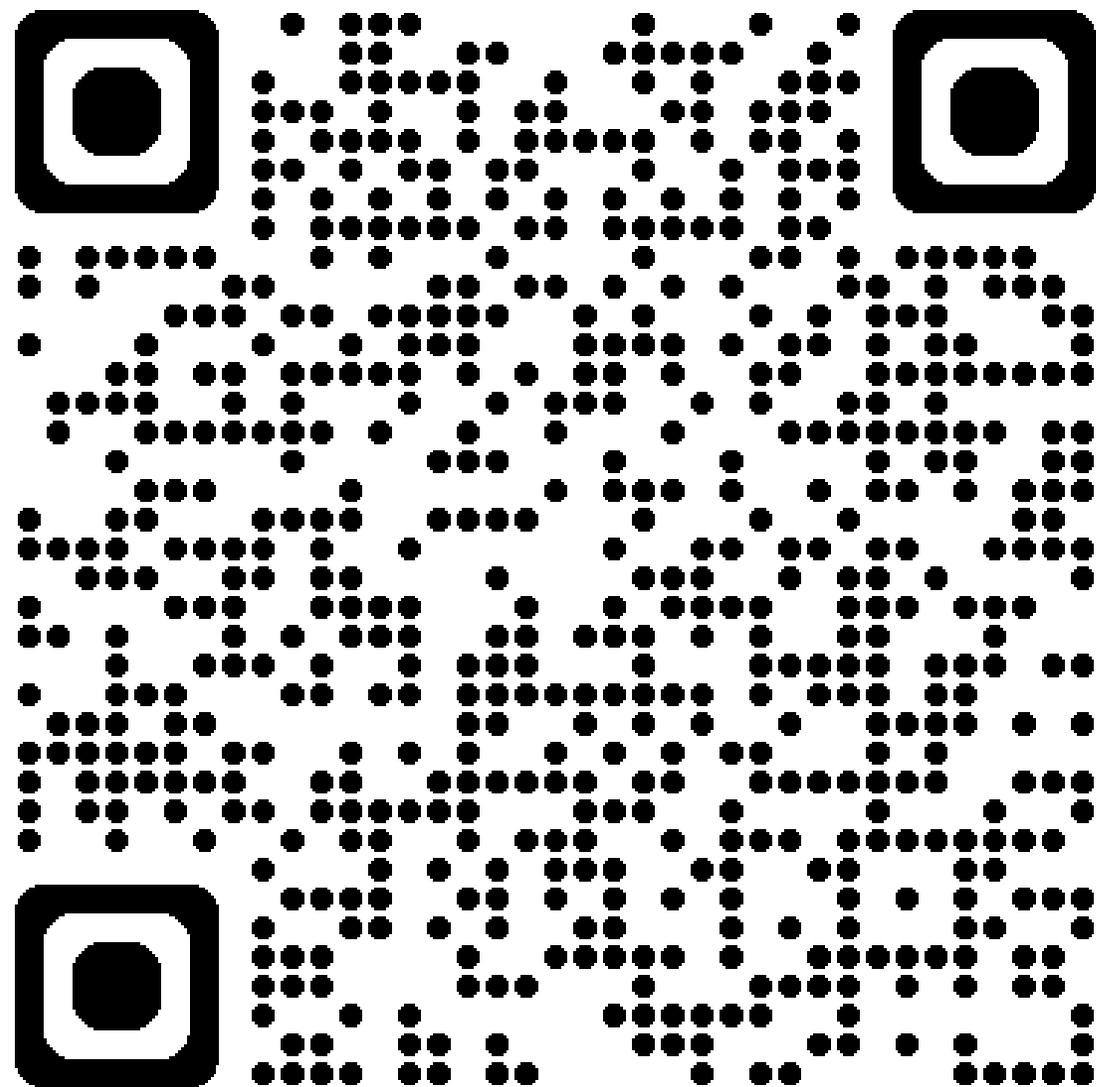
構造学習

構築したネットワークから  
自由に推論が行える  
ネットワーク編集機能



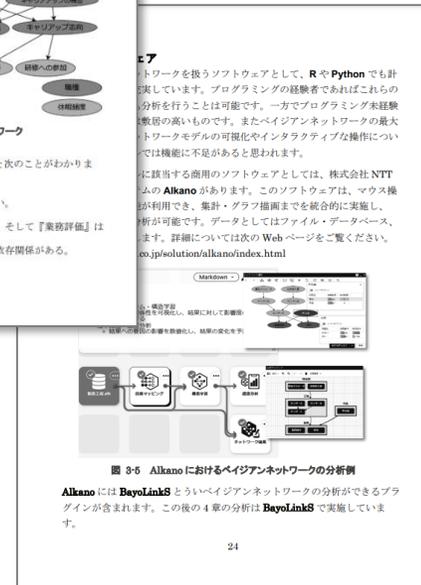
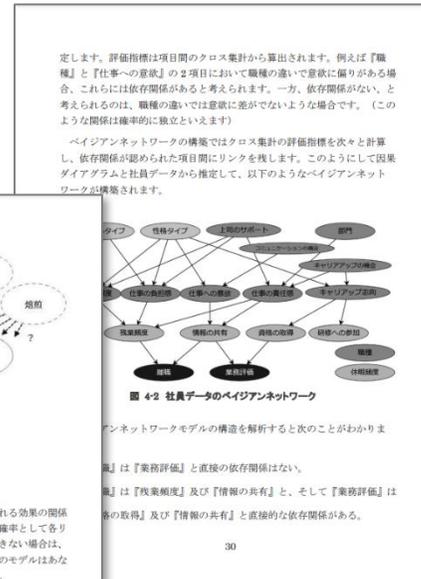
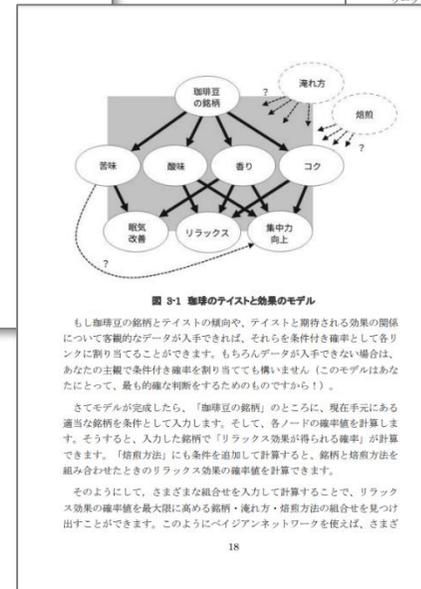
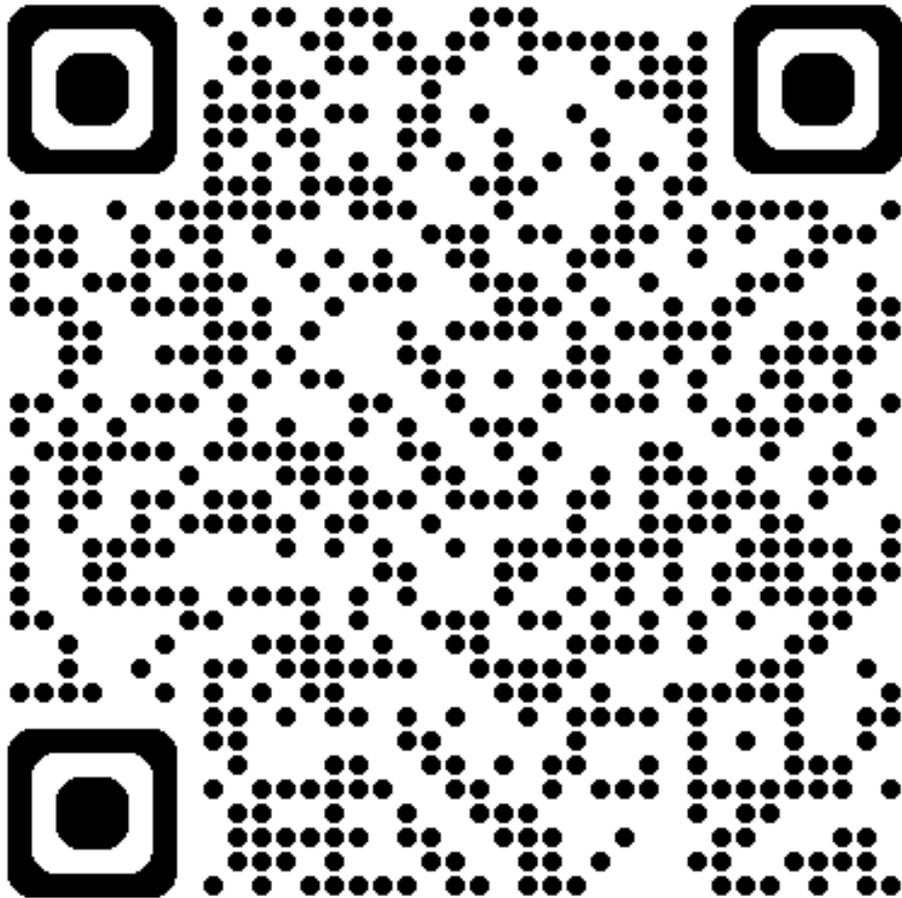
# MSIISM で NTTデータマネジメントサービス様の事例をご紹介します

- NTTデータマネジメントサービス株式会社 様  
社員アンケートデータ分析で  
離職防止施策を見つける事例



<https://www.msiism.jp/article/nttd-ms-wellbeing-at-work.html>

# ご要望にお応えして、 ベイジアンネットワーク入門読本 ダウンロード開始！



<https://www.msiism.jp/ebook.html>

The image features a low-angle, wide shot of a modern city skyline under a clear blue sky. Two prominent skyscrapers with white facades and dark window bands are the central focus. Other buildings of varying heights and architectural styles are visible in the background and foreground. The overall scene is brightly lit, suggesting a clear day. The text 'NTT Data' is superimposed in the center of the image in a bold, white, sans-serif font.

**NTT Data**