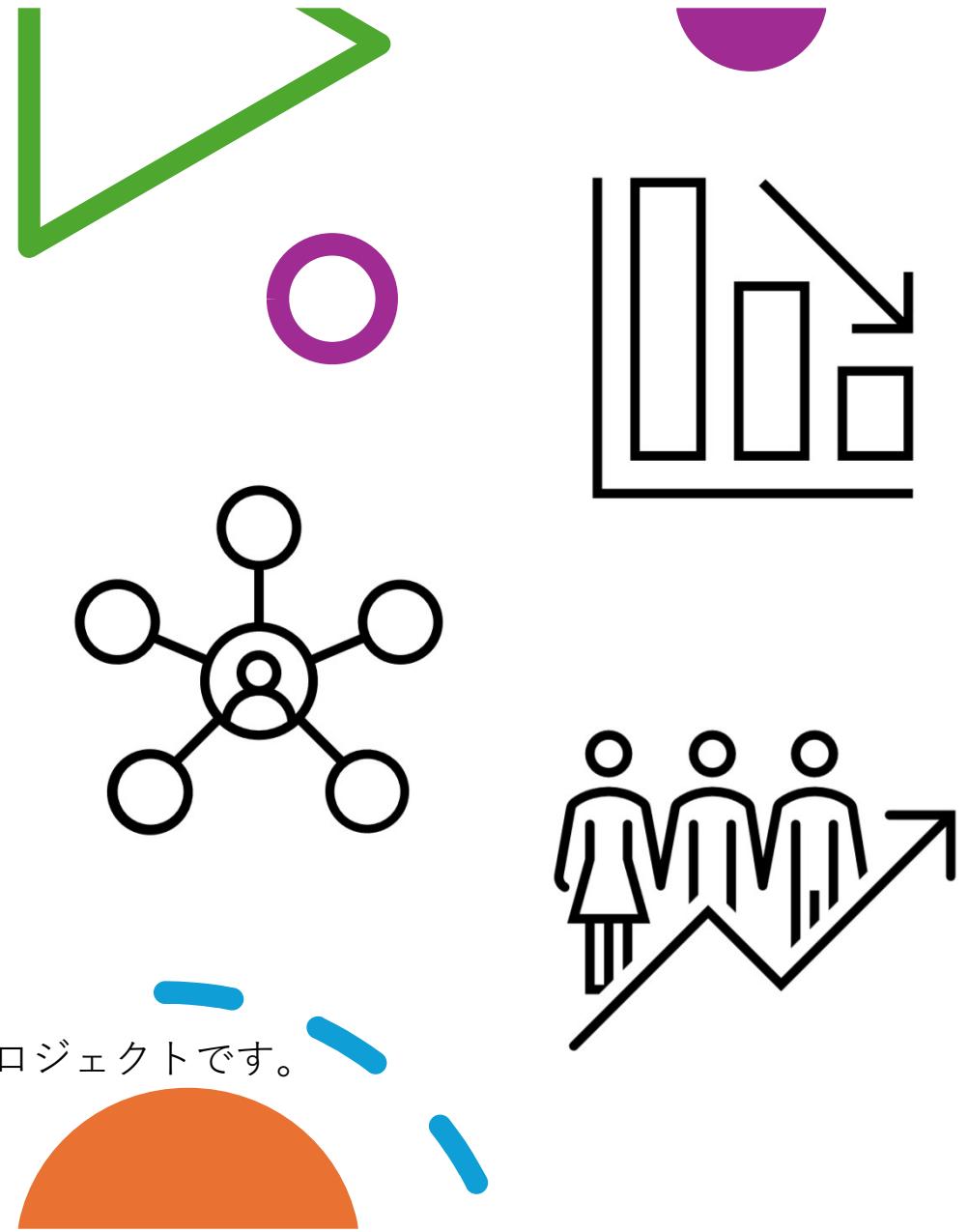


Employee Attrition Analysis (Portfolio Version)

Author : Yuu
Date : 2025

本資料は匿名化されたサンプルデータを用いた個人プロジェクトです。



目 次

I. 導入・背景

- 1. 市場分析
- 2. 課題設定と目的変数の定義

II. データ分析（EDA）

- 3. データ概要と離職率
- 4. 相関ヒートマップ
- 5. 職務満足度別の離職率
- 6. リモートワークとストレス評価と離職率

III. モデル構築準備

- 7. 特徴量エンジニアリングの概要
- 8. 今回の処理ステップ

IV. モデル評価と改善

- 9. モデル構築結果（評価比較）
- 10. 今後の改善案

V. 施策提案と考察

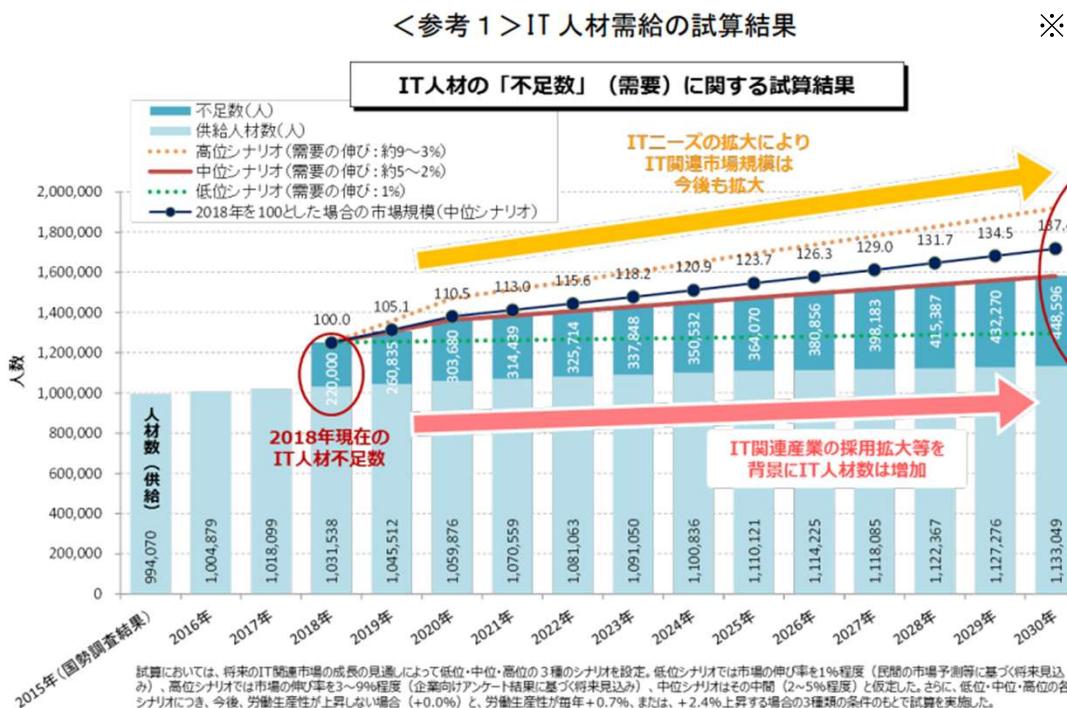
- 11. 特徴量の重要度（ロジスティック回帰）
- 12. 特徴量に基づく施策提案

VI. 成果の見込みと展望

- 13. 留職予測モデルのKPIと期待利益
- 14. お見積り（想定コスト）
- 15. 今後の展望

I. 導入・背景

1. 市場分析



出典 :

- ※1 経済産業省『IT人材需給に関する調査（概要）』
https://www.meti.go.jp/policy/it_policy/jinrai/gaiyou.pdf
- ※2 厚生労働省『IT人材需給に関する調査』
<https://www.mhlw.go.jp/content/12602000/000844586.pdf>

➤ 深刻化するIT人材不足

※1

2018年: 既に約22万人の不足

2030年予測: 最大で約79万人の不足に拡大

➤ IT需要の拡大により人材不足は3.6倍に深刻化 需要動向

- IT需要: 年々増加傾向
- 人材供給: 緩やかな増加にとどまる
- 重要供給のギャップ: 年々拡大する構造

➤ 将来予測の内訳

2030年の不足数:

- 上位シナリオ: 約79万人
- 中位シナリオ: 約45万人
- 下位シナリオ: 約16万人

➤ 離職防止市場の成長と対策

※2

- 離職防止・人材定着を支援する市場が拡大中。
- 「従業員エンゲージメント分析」や「離職予測」などのHRテック導入が進展。
- 離職リスクの可視化により、人的資本の最適化を目指す企業が増加。

I. 導入・背景

2. 課題設定と目的変数の定義

【課題設定】

I社では離職率が16.2%と全産業平均(約12%)を大きく上回っており、人材流出が課題となっていると思われます。そこで本提案では、離職予測モデルを構築することで離職要因を可視化し、離職率の低減と人材定着の実現を目指します。

【目的変数】

「Attrition(離職有無)」を目的変数に設定しました。これにより、どの社員が離職しやすいかを予測し、人事施策の強化に活用します。

※データは約1500名の社員情報を基に分析を行います。

II. データ分析 (EDA)

3. データ概要と離職率

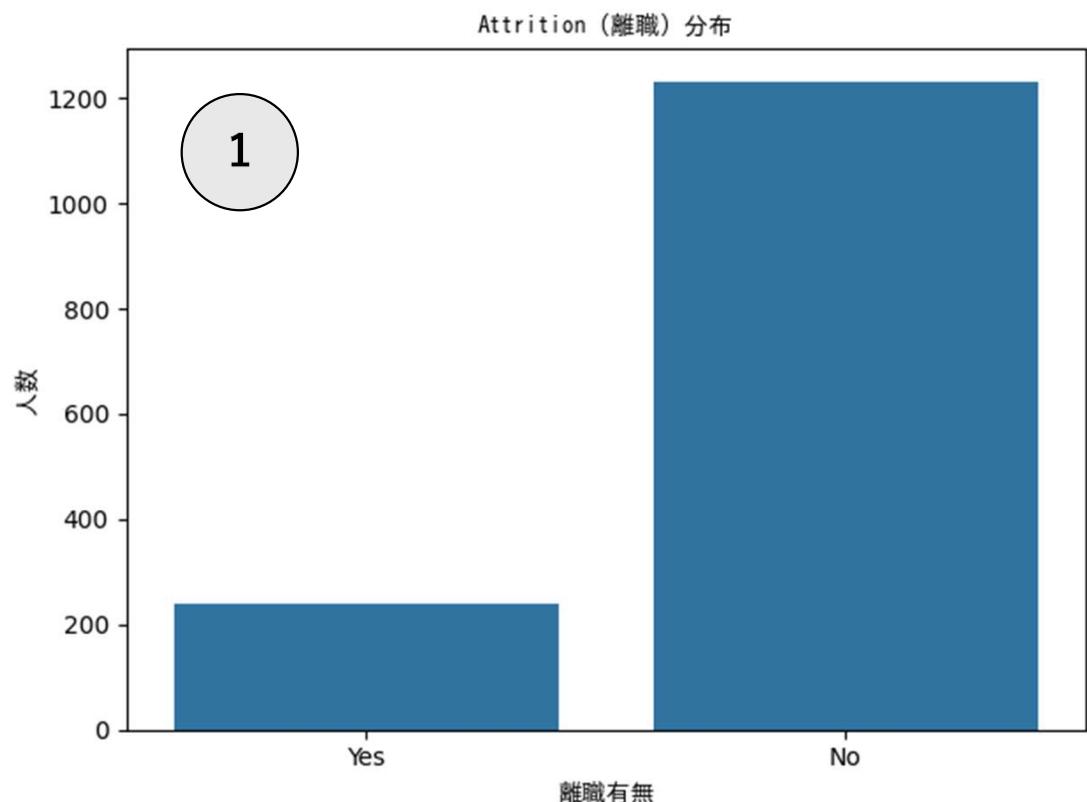
■ データ概要

- ・従業員数: 1470人
- ・特徴量数: 44
- ・欠損値なし

■ 目的変数(離職有無 Attrition)の分布

- ・離職者割合: 約16%
- ・定着者割合: 約84%

→ 約6人に1人が離職しており、
早期発見と対策が必要



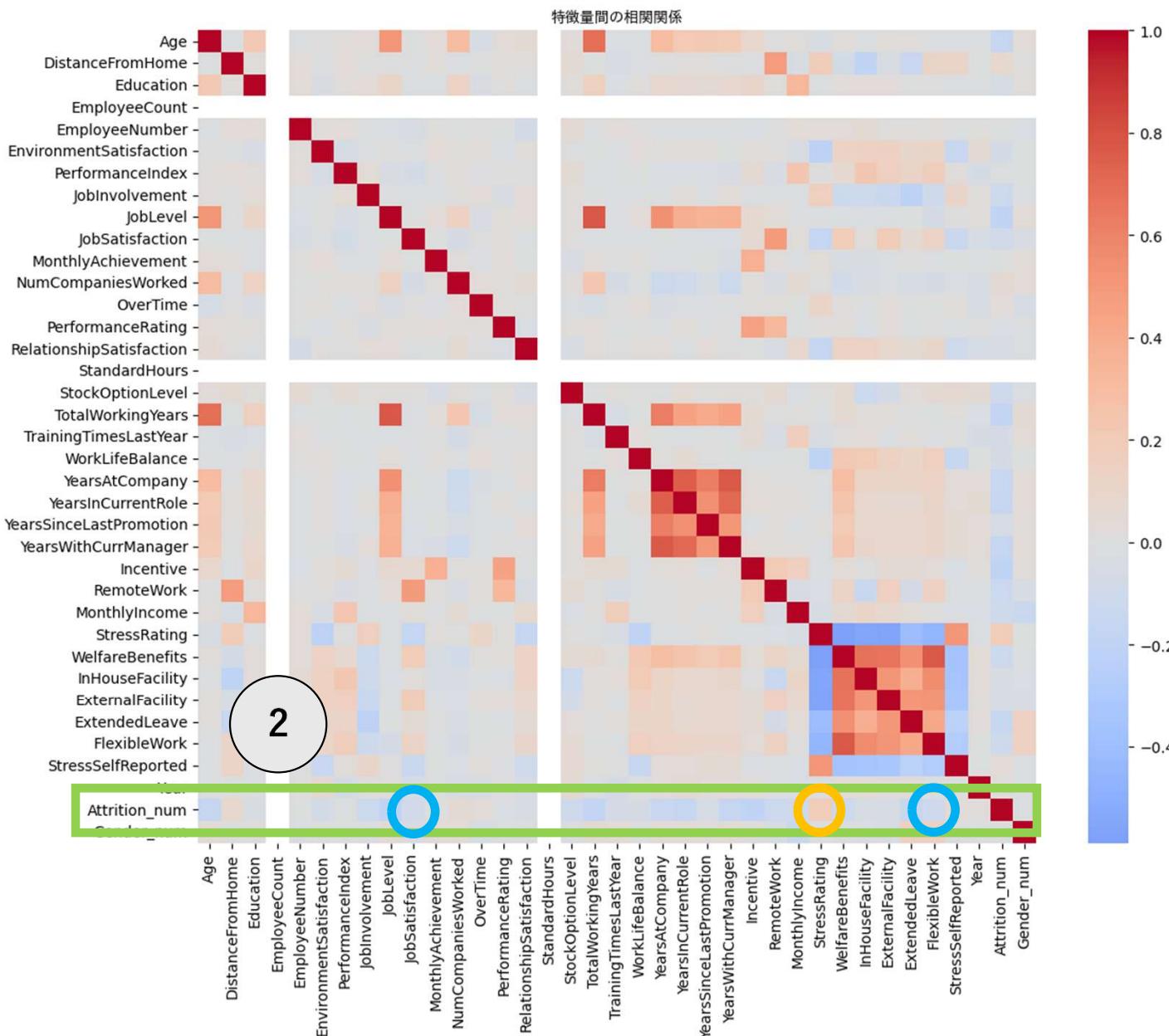
II. データ分析 (EDA)

4. 相関ヒートマップ

■ 数値特徴量間の相関関係

- ・離職有無(Attrition)と
強い関連を持つ変数を特定
- ・例:「ストレス評価」
「職務満足度」
「ワークライフバランス」等

→これら(丸をつけたところ)に
着目し、EDAを試みた。



II. データ分析 (EDA)

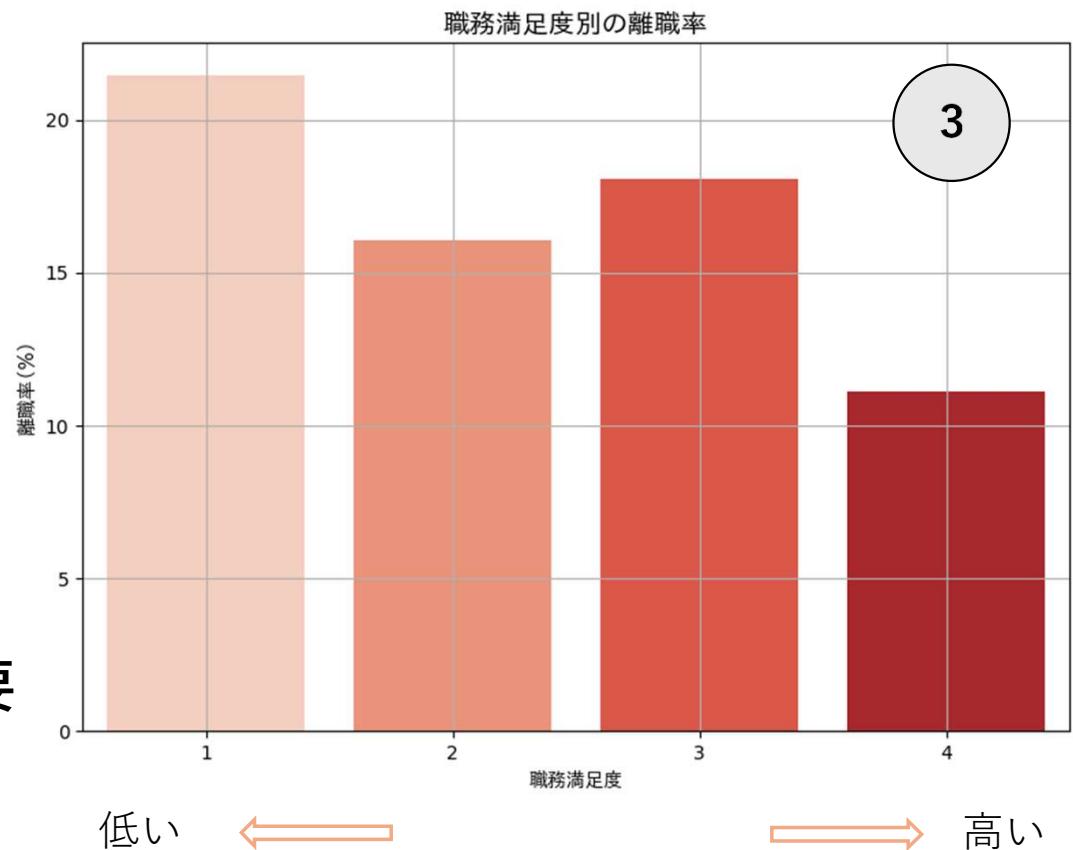
5. 職務満足度別の離職率

・満足度が低いほど離職率が高い傾向が顕著

満足度「1」→約21.5%が離職

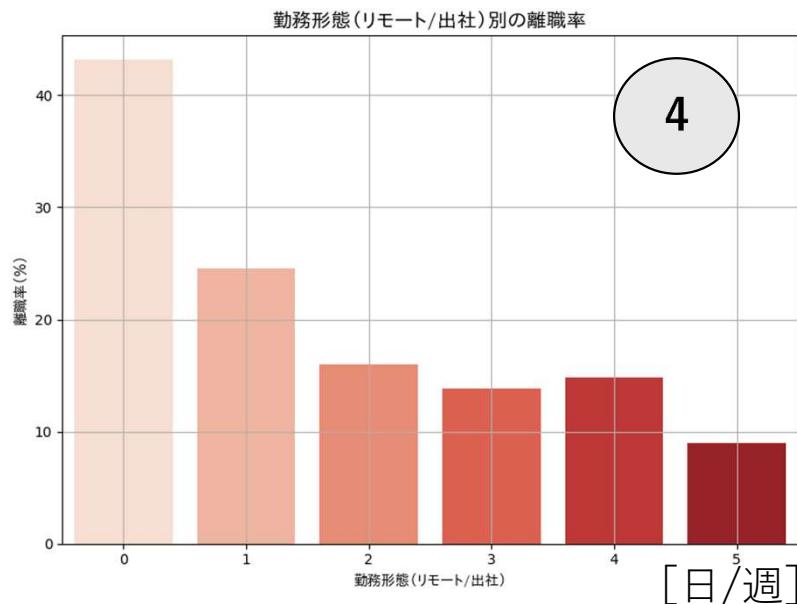
満足度「4」→約11.1% が離職

⇒「職務満足度向上」は離職抑制に重要



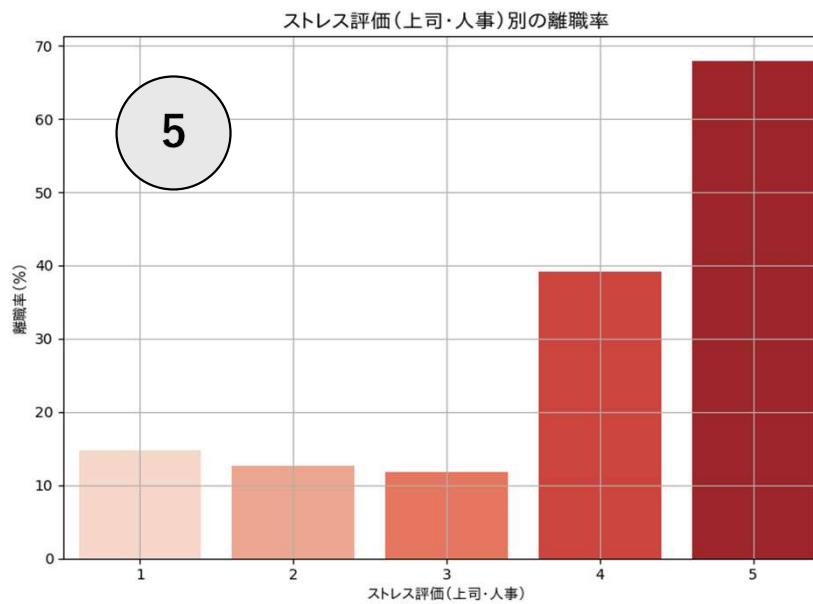
II. データ分析 (EDA)

6.リモートワークとストレス評価と離職率



■ リモートワーク別離職率

出社のみは離職率43%、フルリモートは9%
柔軟な勤務形態が鍵。



■ ストレス評価別離職率

ストレスが高い社員ほど離職率が急増(67.9%)

⇒ リモート勤務やストレス軽減策の導入が有効と推察

III. モデル構築準備

7. 特徴量エンジニアリングの概要

- **特徴量エンジニアリングとは？**

モデルの性能を最大化するため、
データを変換・加工して特徴量を
整理するプロセス

- 精度向上のために、右の内容を
実施する

処理内容	目的
不要カラムの削除	ノイズや意味のない情報を排除
ラベルエンコーディング	2値カテゴリを数値に変換
One-Hotエンコーディング	多カテゴリをダミー変数に変換
特徴量の作成	ドメイン知識を活用して意味のある変数を追加
スケーリング	モデル学習を安定させるため正規化

III. モデル構築準備

8. 今回の処理ステップ

モデルの性能を最大化するため、以下の処理を実施：

- **不要な列の削除**：一意 or 定数（例：EmployeeNumberなど）
- **カテゴリの数値変換**：LabelEncoder（2値） / OneHotEncoder（多値）
- **新しい特徴量の作成**：勤続年数比（TenureRatio）、ストレス指標の平均など
- **スケーリング**：月収の正規化でバイアス防止
- 处理後、特徴量数は**約44→60項目に拡張**され、モデル精度の向上に貢献

IV. モデル評価と改善

9_1. モデル構築結果(評価比較)

<モデル評価指標（抜粋）>

モデル名	Accuracy 正解率	Recall 再現率	Precision 適合率	F1-score 再現率と適合率のバランス	AUC 総合力
LogisticRegression ロジスティック回帰	0.88	0.42	0.74	0.53	0.82
RandomForest ランダムフォレスト	0.85	0.15	0.78	0.25	0.82
XGBoost	0.85	0.29	0.61	0.39	0.80

IV. モデル評価と改善

9_2. モデル構築結果(評価比較)

【考察】

- ・ロジスティック回帰がRecall（離職検出率）とAUCともに最も高く、安定した性能を発揮
- ・ランダムフォレストは精度は高いが離職をほぼ見逃しており実務では使いづらい
- ・XGBoostは精度バランスが良いがRecallがやや低め

離職予測では「**Recall重視**」が基本 ⇒ 離職者を検出できなければ事前対応が困難に

【結論】

- ・ロジスティック回帰を最終モデルに採用
- ・精度・説明性・実務利用（人事への説明のしやすさ）も加味

～今後の活用～

- ・モデルの学習・予測→離職リスクのスコア化に活用
- ・特徴量重要度の分析→離職に強く関係する変数の特定
- ・実ビジネス施策への接続→ストレスの高い社員への支援策の設計

IV. モデル評価と改善

10. 今後の改善案

改善ポイント	内容例
特徴量最適化	相関が高すぎる変数の除去、変数の選択的活用
サンプリング手法	SMOTEなどで 離職データを補完し、バランス化
閾値調整	0.5以外のしきい値で Recall優先の最適化
ハイパーパラメータ調整	GridSearchCVやOptunaなどで モデル性能を最大化
アンサンブル構成	モデル融合で 精度とRecallのバランスを向上させる

V. 施策提案と考察

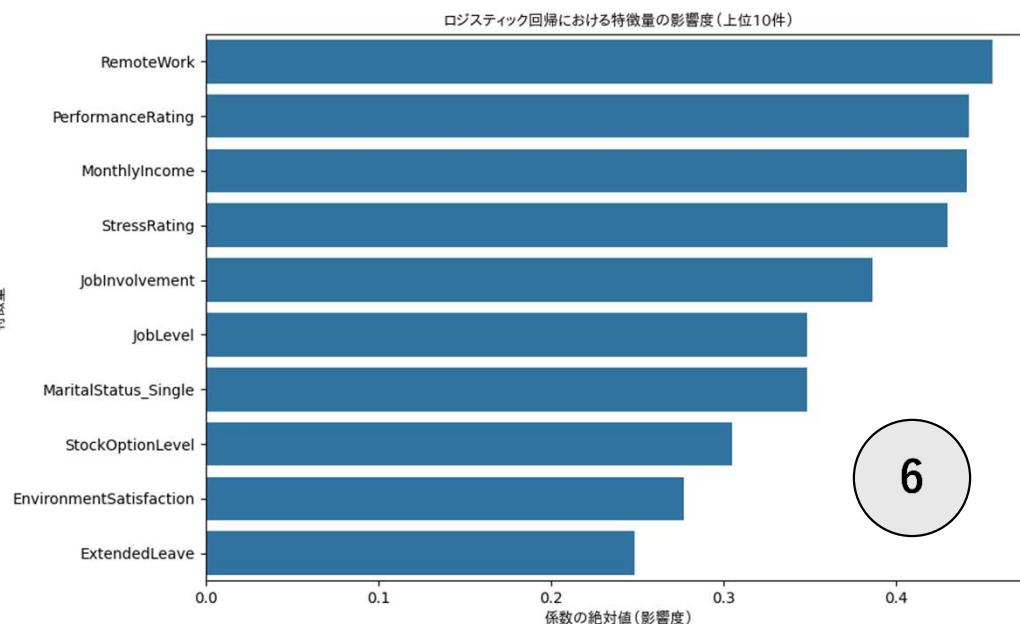
11. 特徴量の重要度(ロジスティック回帰)

特徴量重要度(上位10)

係数の絶対値が大きい順(モデルに影響を与える順)

AUC: 0.82 と比較的良好なモデル精度。

主に「働き方・報酬・満足度」が重要な決定因子と読み取れる。



順位	特徴量	係数の符号	解釈(符号の意味)
1	RemoteWork (頻度)	-	リモート頻度が高いほど離職しにくい
2	PerformanceRating	-	高評価者ほど離職しにくい
3	MonthlyIncome (月収)	-	月収が高いと離職しにくい
4	StressRating (ストレス)	+	ストレスが高いと離職リスクが上がる
5	JobInvolvement (仕事関与)	-	関与度が高いと離職リスクが下がる
6	JobLevel (職位)	-	職位が高いと離職しにくい
7	MaritalStatus_Single	+	独身の方が離職リスクが高め
8	StockOptionLevel	-	ストックオプションがあると定着率が高い
9	EnvironmentSatisfaction	-	環境満足度が高いと離職しにくい
10	ExtendedLeave (特別休暇)	-	休暇制度があることで定着が促進される

V. 施策提案と考察

12. 特徴量に基づく施策提案

施策	実施主体	実施時期	実施内容
リモート勤務拡充	人事部 + 管理職	2025年10月～	原則週2回以上の在宅勤務を許可
ストレスケア強化	産業医 + メンタル支援チーム	2025年11月～	定期メンタルチェック + 業務量調整面談
報酬制度の見直し	人事戦略室	2026年1月～	高評価者への昇給 + 役職基準の見直し
独身者支援策	福利厚生チーム	2026年4月～	社内交流会・住宅補助の導入検討

<ポイント>

離職リスクの可視化と早期対応が可能になる。

特に「ストレス・働き方の柔軟性・報酬制度・満足度」が離職率に大きく影響。

よって社員一律ではなく、個別最適化された対応が重要であり、モデルの活用により離職率の定量的な改善も期待できる。

VI. 成果の見込みと展望

13. 離職予測モデルのKPIと期待利益

■ 現状データ（年間）

- ・ 社員数：1,470人
- ・ 離職率：16.19%
→ 離職者数： $1,470 \text{人} \times 16.19\% = 238\text{人}$
- ・ 1人あたりの離職コスト：約300万円

■ 改善目標

- ・ 離職率：12%
→ 離職者数： $1,470 \text{人} \times 12\% = 176\text{人}$
- ・ 改善による削減人数： $238\text{人} - 176\text{人} = 62\text{人}$

■ 期待利益の試算

- ・ 削減によるコスト削減：
 $62\text{人} \times 300\text{万円} = \underline{\text{1億8,600万円}}$

指標	現状	目標	差分
離職率	16.19%	12%	-4.19pt
年間離職者数	238人	176人	-62人
離職コスト(試算)	約7.1億円	約5.3億円	-1.86億円

■ 組織への効果

- ・ 人材流出の抑制
- ・ 採用、教育コストの軽減
- ・ チームの生産性・安定性の向上

VI. 成果の見込みと展望

14. お見積り(想定コスト)

【期待効果】年間、約1.86 億円のコスト削減(離職減による)

【コスト対効果】初年度 ROI 約 37 倍

項目	金額(年額)	備考
モデル構築・初期導入費	約 200 万円	コンサル費・初期開発費含む
システム連携・ダッシュボード構築	約 150 万円	BIツール・社内人事DB連携など
社内運用トレーニング	約 50 万円	人事担当者向け研修等
メンテナンス・改善運用費	年間 約 100 万円	モデル更新・チューニング費
合計(初年度)	約 500 万円	(税抜き・参考価格)

VI. 成果の見込みと展望

15. 今後の展望

◆ 繼続的なモデルアップデート

「 新たなデータ（社内アンケート・勤怠ログ等）の取り込み」

◆ リアルタイムな離職リスクアラートの構築

「 ダッシュボード化・個別通知による早期介入」

◆ 生成AIとの連携による人事アドバイスの自動化

「 チャットボットが相談対応・キャリア設計補助など」

Note : This report and all visualizations were created by Yuu as a personal portfolio project. All data used is synthetic or anonymized.

最終目的：データに基づく「人が辞めない組織づくり」