

# 機械学習を活用した従業員離職予測の導入提案

従業員の離職リスクを予測し、企業の人事戦略を最適化

松澤 建門

# 労働市場での離職率による影響

近年の労働市場では、特に若年層や低教育レベルの従業員が早期に離職する傾向が強いことが分かっている<sup>(1)</sup>。

従業員の離職率が高いと、企業の業績に悪影響を及ぼす可能性がある。Maertz<sup>(2)</sup>らは若年層の離職率が高い場合、企業のパフォーマンスにどのような影響があるのかを調査した。若年層の離職は短期では採用コストや教育コストがかかるだけだが、長期的には企業の競争力やイノベーションを弱体化させる可能性が高いことを示した。特に、技術革新が必要な業界では企業の成長に悪影響を及ぼすと述べられている。

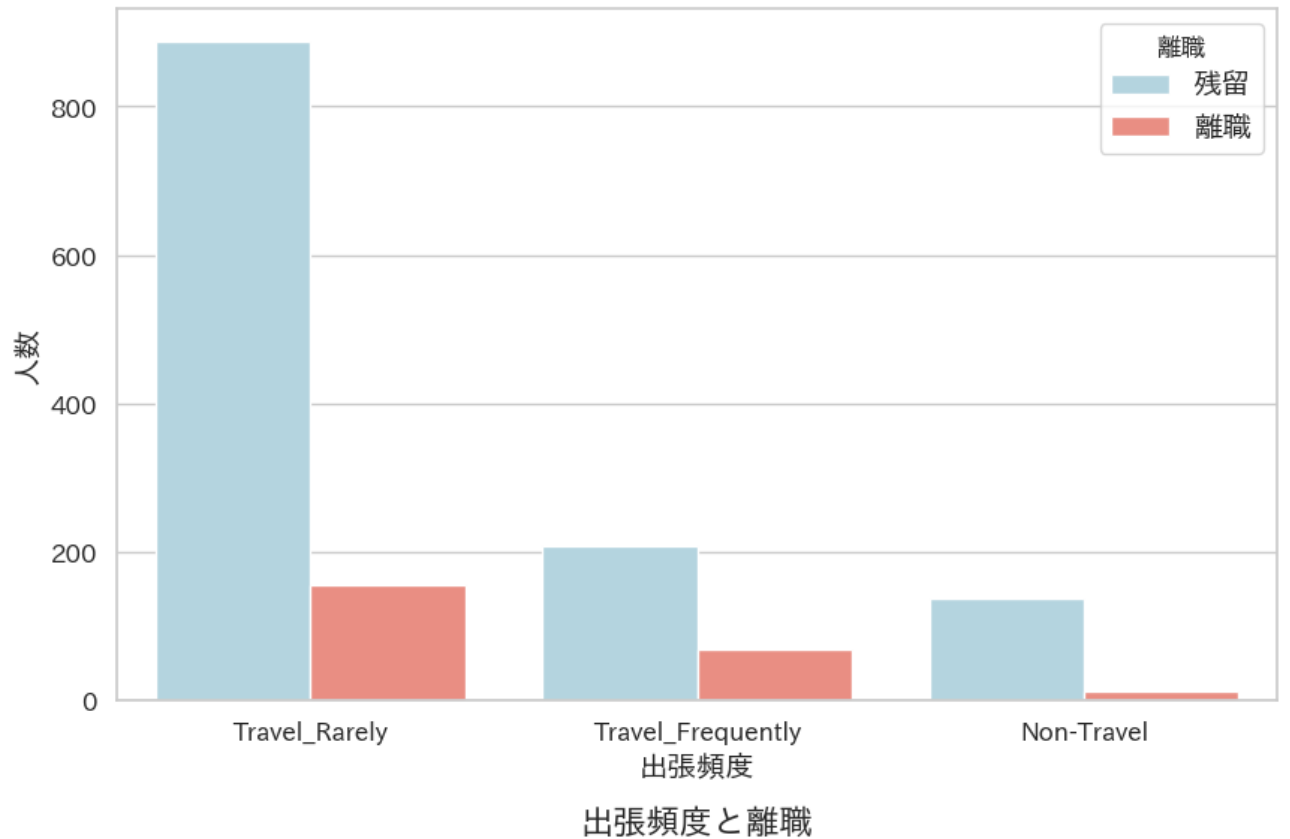
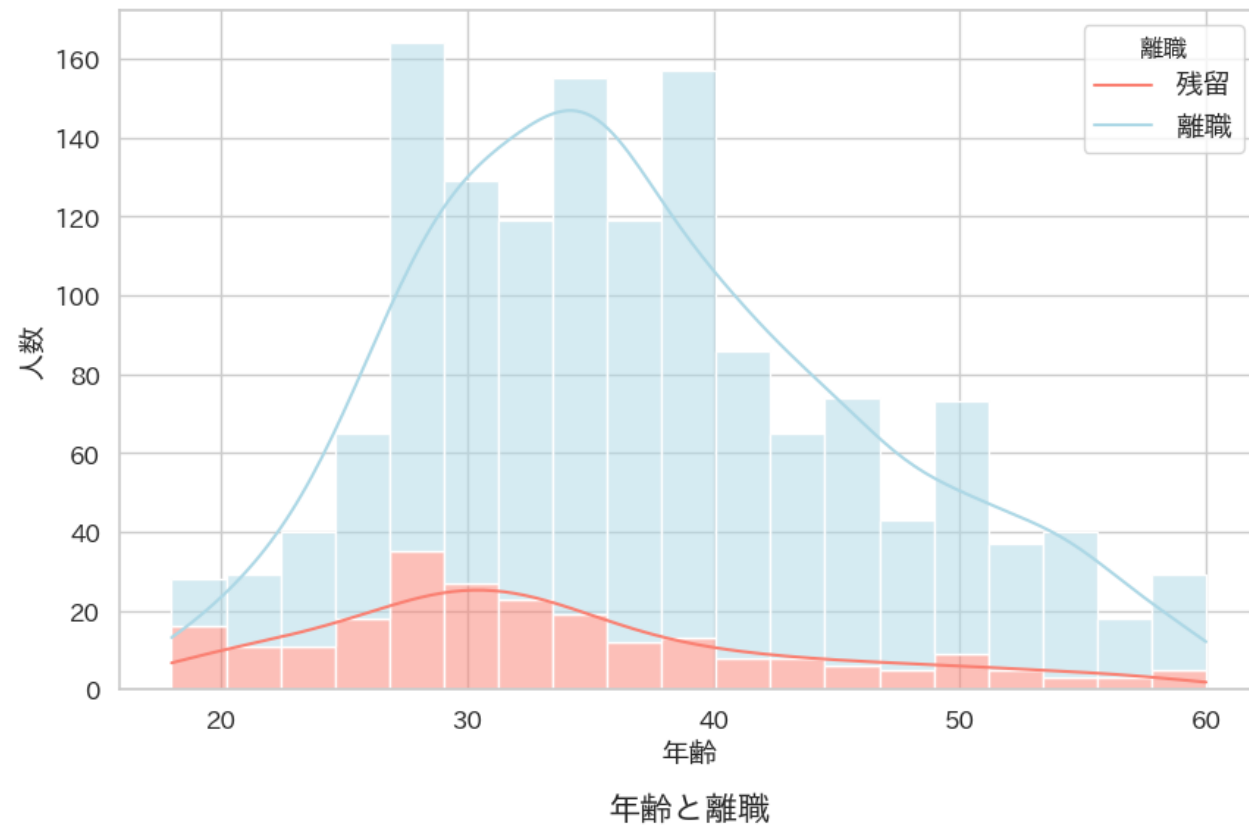


I社でも同様に若年層の離職が高く (3ページの左図参照) 対策が必要である

出典(1)：若手人材の早期離職の実態～，[https://career-research.mynavi.jp/column/20230612\\_52469/](https://career-research.mynavi.jp/column/20230612_52469/)，閲覧日(2025年1月19)

出典(2)：Maertz(2007), The Effects of Perceived Organizational Support and Perceived Supervisor Support on Employee Turnover, Journal of Organizational Behavior 28(8):1059-1075

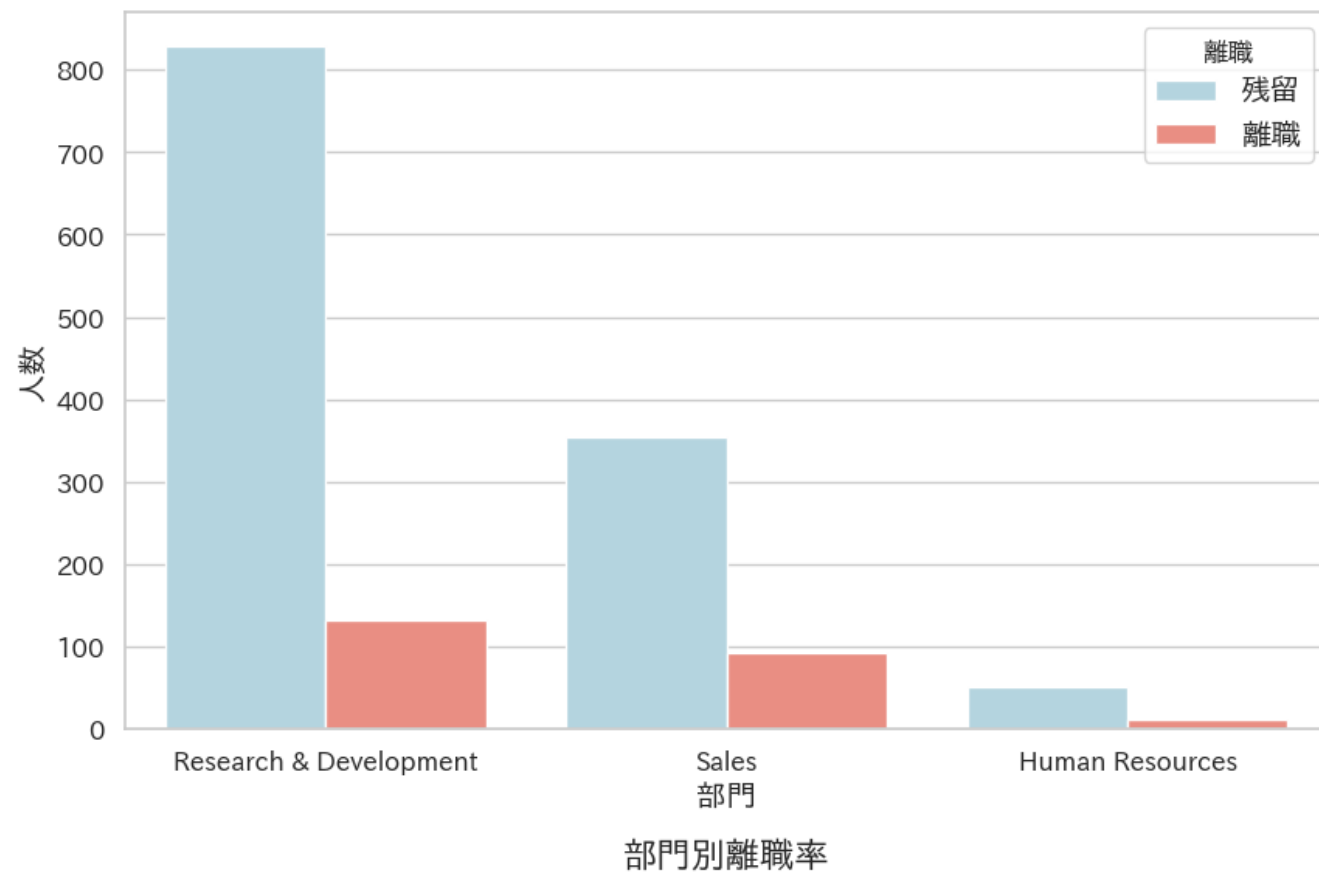
# EDA可視化



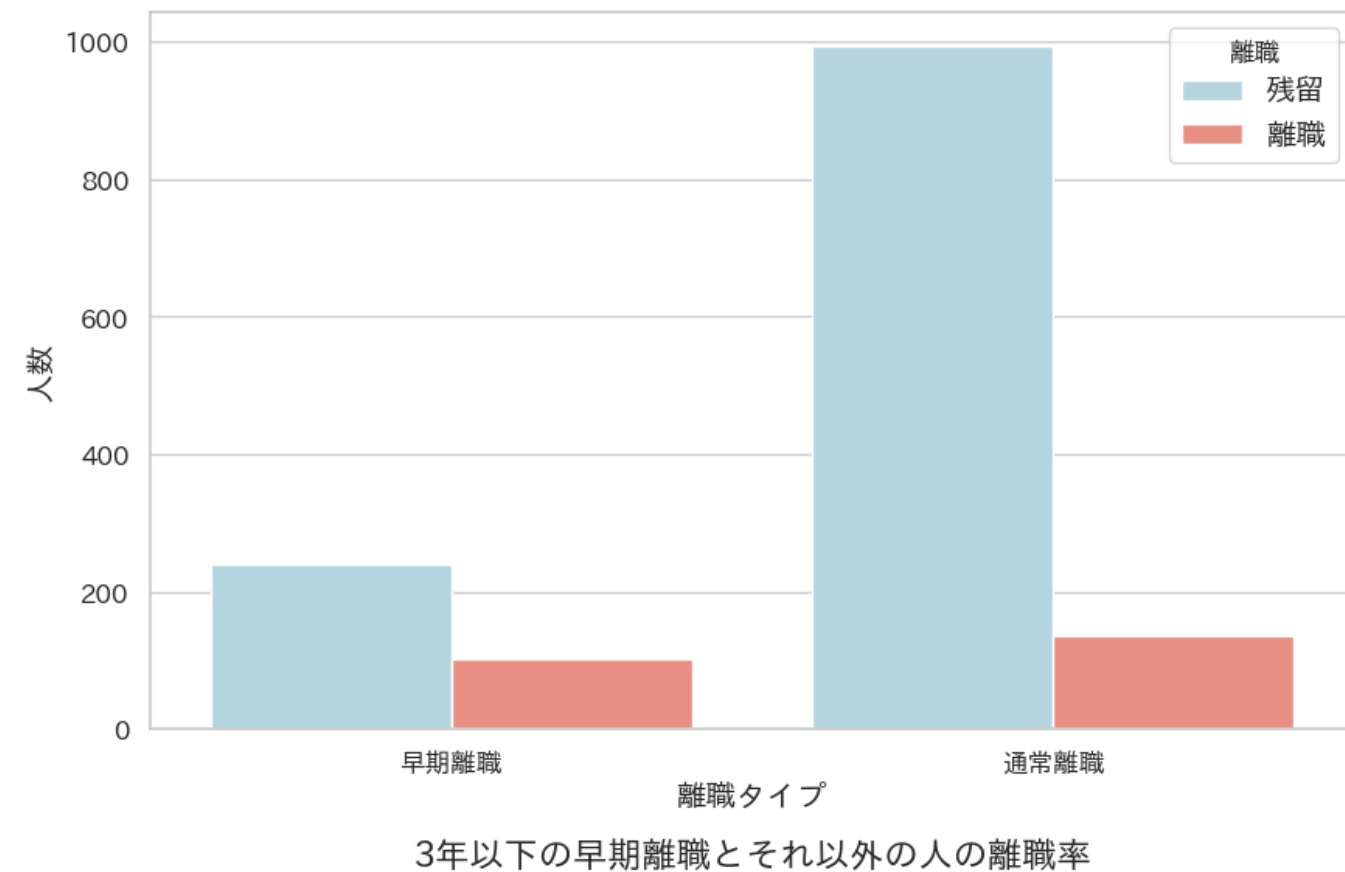
若い従業員（特に20代から30代前半）の離職率が高いことがわかる。若い世代はキャリアの選択肢が多く、特に職場の文化が合わないと感じた場合、転職を選びやすい傾向にあると考えられる。逆に、年齢が上がるにつれて離職率が低くなるのは、安定を求める傾向が強くなるため、転職を避ける傾向があるからだと考えられる。

出張頻度が高い従業員において離職率が高い傾向がわかる。頻繁な出張は身体的な負担となり、生活の質を下げる要因となるため、転職を選択する要因となる。出張が少ない従業員は、離職率が低く、安定して働きやすい環境にあると考えられる。

# EDA可視化



営業部門や人事部門は、業務上のプレッシャーが高く、達成しなければならない数字上の目標が多いと考えられる。そのためストレスや仕事の負担が増加し、それが離職につながっていると考えられる。



3年以上在籍している従業員は、会社の環境に適合しており、在籍が3年未満の人と比べると離職率が低くなっていると考ええる。そのため、入社3年未満の人に重点してケアが必要であることがわかる。

# 予測に用いた手法

データを入力

カラム名
Age
Attrition
DailyAchievement
Department
DistanceFromHome

・  
・  
・  
・

頂いたデータ以外に新たな  
特徴量を作成  
カテゴリ変数を数値化

LightGBM,XGBoost,  
ロジスティック回帰を使用

データの  
前処理

予測モデ  
ルの作成

予測結果  
を出力

機械学習のモデル  
を作成するために  
重要な過程

# 予測に用いた手法(特徴量の作成)

- なぜ、新たな特徴量を作成する必要があるの？

1. 隠れた関連性を引き出せる：データの中には、単体では影響が小さい特徴量でも、他の特徴量と掛け合わせることで予測に重要な情報が浮き彫りになる場合があるため
2. モデルの予測精度向上：新たな特徴量はモデルが学習するため、より多くの情報を提供可能であるため

- 新たに作成した特徴量

- JobSatisfaction x EnvironmentSatisfaction: 仕事満足度と職場環境の満足度を掛け合わせた。個別の満足度では見えない総合的な満足感が離職率に与える影響を把握できることを期待
- TotalWorkingYears x Education: 仕事の経験年数と教育レベルを掛け合わせることで、従業員が持つ専門的な経験に関する情報を抽出。教育と経験の両方が高い従業員がどのように振る舞うのかに着目
- PerformanceRating x Incentive: 従業員のパフォーマンス評価とそのインセンティブが掛け合わされることで、パフォーマンスに見合った報酬がある場合、どの程度従業員のモチベーションや離職に影響を与えるかに着目
- DistanceFromHome x WorkLifeBalance: 自宅からの距離と仕事のバランスの関係を組み合わせます。長距離通勤が従業員の仕事のバランスにどのように影響するかを捉える

# 予測に用いた手法(モデル)

- ・今回用いた予測モデルは以下の4つ

1. ロジスティック回帰
2. 決定木
3. ランダムフォレスト
4. LightGBM

- ・この4つを選択した理由

1. ロジスティック回帰：特徴量と離職リスクとの関係が単純である場合に効果的
2. 決定木：特徴量の重要度（どの要素が離職に影響しているか）が分かりやすい
3. ランダムフォレスト：「満足度」「収入」「勤続年数」など複数の要因が絡む複雑なデータを扱う場合に効果的
4. LightGBM：「満足度」「収入」「距離」など、多数の特徴量が絡み合うデータセットで効果を発揮

# 予測に用いた手法

以下の2つの前処理を行った

1. 不均衡データの対策
2. 交差検証

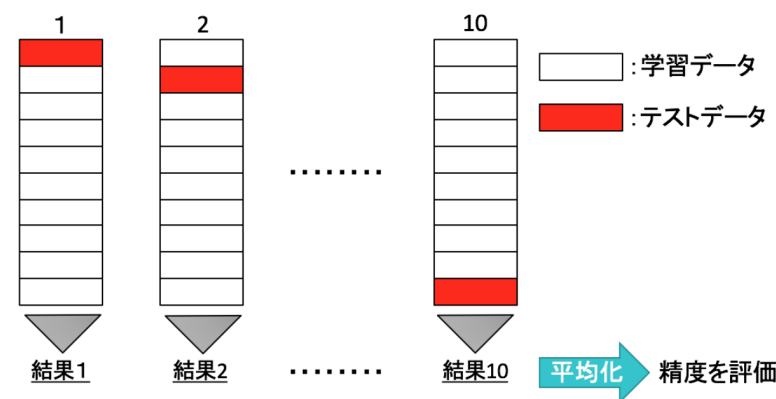
## ・不均衡データとは？

データセット内で特定のクラス（例えば、離職者と非離職者）に属するデータが非常に少ない場合のことを指す。今回の頂いたデータは1471人中、離職者の数が237人と離職者の数が大幅に少ない。

そのため、そのまま機械学習モデルを訓練すると、予測モデルは多数派のデータばかりを覚えてしまい、少数派を無視してしまう可能性がある。

そこで、クラス重みを用いて少数派である離職者のデータに対しても重要だと伝えるために、重みを増やした。これにより、モデルが離職者に対してもっと注目し、適切に予測を行えることが可能。

## ・交差検証とは？



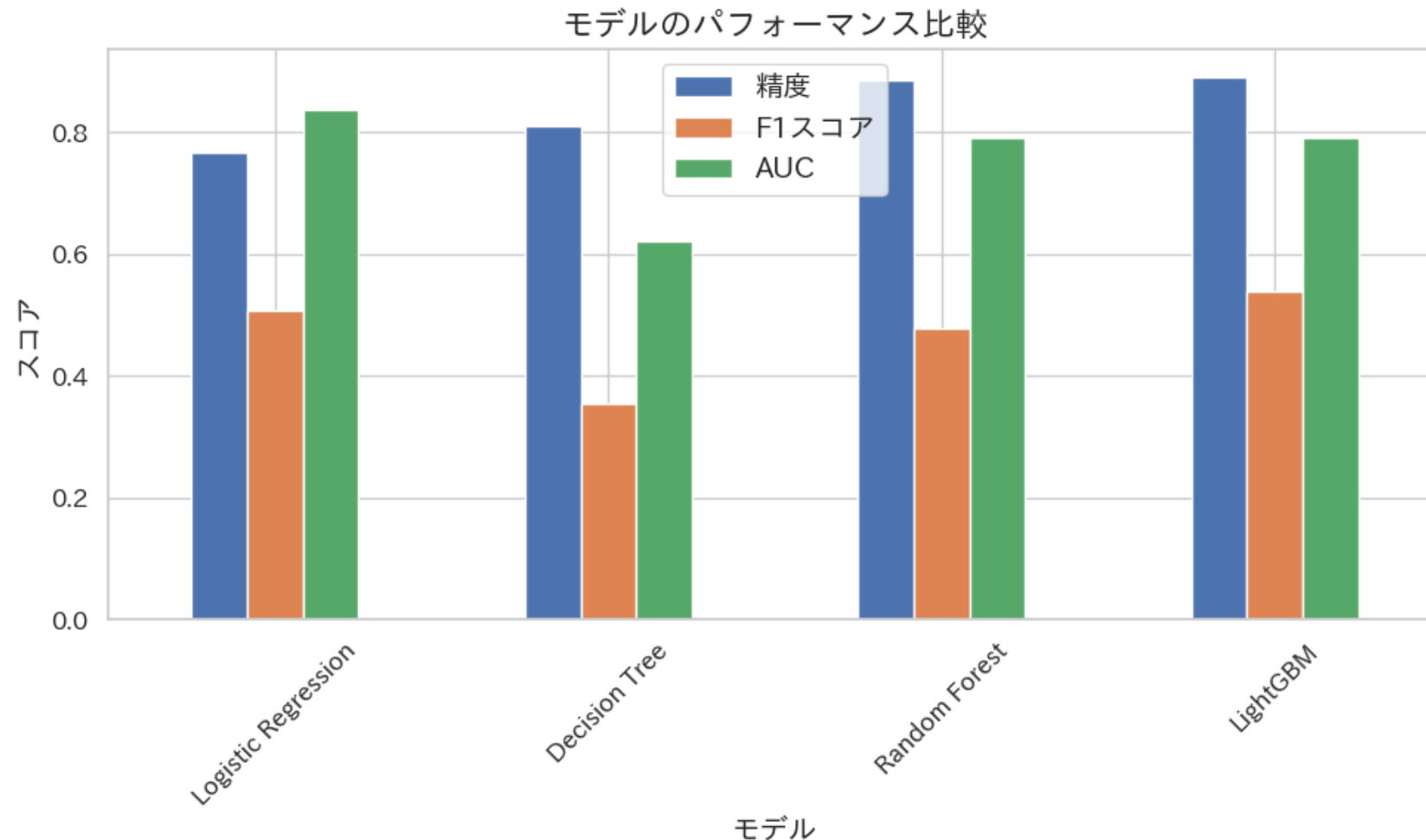
交差検証<sup>(3)</sup>

交差検証は、機械学習のモデルを評価するための手法。モデルの予測精度をしっかりと測るために、データをいくつかの部分に分けて、複数回学習と評価を行う方法である。左の図のように1つのテストデータとして扱い残りの部分を学習データとして扱う。

今回のように、少数派を予測したい場合に交差検証を用いることでモデルの安定性の確認やより信頼性の高い結果を得ることが可能。



# 推定モデルの結果



- ・ 精度：全体の中でどの程度予測できているかの割合
- ・ F1スコア：予測の正確さと見逃しの少なさのバランスを見る指標
- ・ AUC：モデルがどの程度優れているのかを示す指標

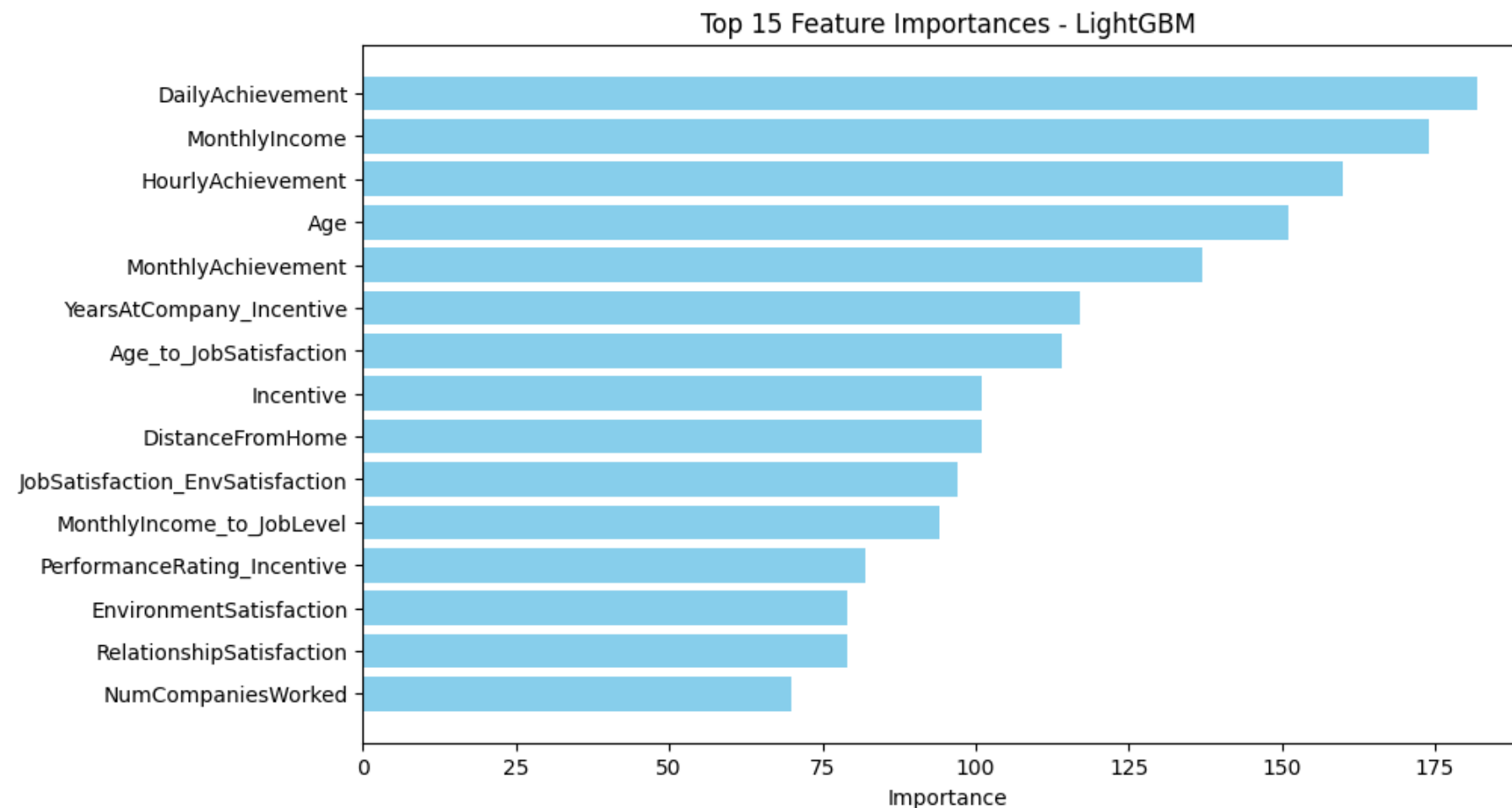
F1スコアとAUCを見ることで、少数派クラス（離職する従業員）をしっかりと予測できているかを判断できる



ランダムフォレストやLightGBMは、AUCやF1スコアで優れた結果を示しており、特に離職した従業員の予測に強いモデルと言える。

特に、LightGBMは精度:0.89, F1スコア 0.54, AUC: 0.79と比較的高いと言える。これにより十分、離職予測モデルとして機能していると言える。

# LightGBMの上位15個の重要な特徴量



従業員の離職予測において、業績やインセンティブ、仕事の満足度、環境の満足度などが最も影響力のある特徴量であることを示している。

# 提案1 (離職予測モデルを活用した早期予測システム)

- ・ 提案内容

I社の従業員データを活用し、離職リスクが高い従業員を早期に特定するための予測モデルを構築する。  
予測結果に基づき、リスクの高い従業員に対して、早期の面談やカウンセリング、役割変更などの介入策を講じることで、離職を防ぐ。

- ・ 効果

- 従業員の離職率の低減: 早期の介入により、リスクの高い従業員を適切にサポートし、離職を防ぐことが可能。これにより、離職率を低減させ、人的資源の安定化を図ることが可能。
- コスト削減: 離職に伴う採用コストや採用外部コストの削減が期待される。採用外部コストは求人広告費用や会社説明会などの運営費用に該当する。また、従業員の業務効率向上にも繋がり、全体的なコストパフォーマンスの向上が見込める。

頂いたデータから退職者は237人いる。ここから離職率が20%削減できると仮定する。これにより47人の離職を防ぐことが可能。

採用活動のコスト:新たに1名の従業員を採用するのに5000人以上の従業員規模の会社では78.5万円<sup>(4)</sup>必要

$$47人 * 78.5 = 3689.5 \text{ 万円}$$

採用外部コスト : ITエンジニアを採用するのに求人広告費用や会社説明会などの運営費用は1名あたり79.7万円<sup>(4)</sup>必要

$$47人 * 79.7 = 3745.9 \text{ 万円}$$

合計 : 7435.4万円コスト削減

# 提案2(従業員のエンゲージメント向上のための分析基盤)

## ・ 提案内容

従業員満足度やエンゲージメントを向上させるための要因分析を行い、離職防止に向けた施策を提案する。  
機械学習を用いて、従業員のエンゲージメントに強く影響する要因を特定し、改善策を提示する。  
例えば、給与、ワークライフバランス、インセンティブ、仕事の満足度、環境の満足度など離職を引き起こす要因に対応するための最適な戦略を提示する。

## ・ 効果

- 従業員の満足度とエンゲージメント向上: エンゲージメント向上に必要な施策を特定し、従業員のモチベーションを高めることが可能。従業員の満足度が高いほど、離職のリスクは低くなる。
- 組織文化の改善: エンゲージメントの向上は、職場の文化や雰囲気の改善に繋がり、企業全体のパフォーマンス向上に繋がる。
- コスト削減: 提案1と同様に離職者を削減可能。離職者が提案1よりさらに2%抑制できると仮定すると5人追加で離職を防ぐことが可能。

採用活動のコスト:新たに1人の従業員を採用するのに5000人以上の従業員規模の会社では78.5万円<sup>(4)</sup>必要

$$5人 * 78.5 = 392.5 \text{万円}$$

採用外部コスト:ITエンジニアを採用するのに求人広告費用や会社説明会などの運営費用は1人あたり79.7万円<sup>(4)</sup>必要

$$5人 * 79.7 = 398.5 \text{万円}$$

パフォーマンス向上:高いエンゲージメントにより生産効率が向上することが期待できる。仮に業務効率が5%向上すれば年間売上高1億円(仮)のI社では500万円の増益が期待可能。

利益向上可能合計:1291万円

# 投資収益率

- ・ 離職予測モデル（提案概要 1）：

- 初期投資: 予測モデルの開発費用（データ準備、モデル構築、評価、運用）は1000万円年間
- コスト削減: 約7435万円のコスト削減が見込まれるため、 $ROI = (7435万円 - 1000万円) / 1000万円 =$  約6.4倍の効果が期待可能。

- ・ 従業員エンゲージメント向上（提案概要 2）：

- 初期投資: エンゲージメント分析基盤の構築やデータ収集、システム導入には200万円年間
- 利益増加: 1291万円の利益増加が見込まれるため、 $ROI = (1291万円 - 200万円) / 200万円 =$  約5.5倍の効果が期待可能。