

|社における

離職者予測モデルのご提案



目次

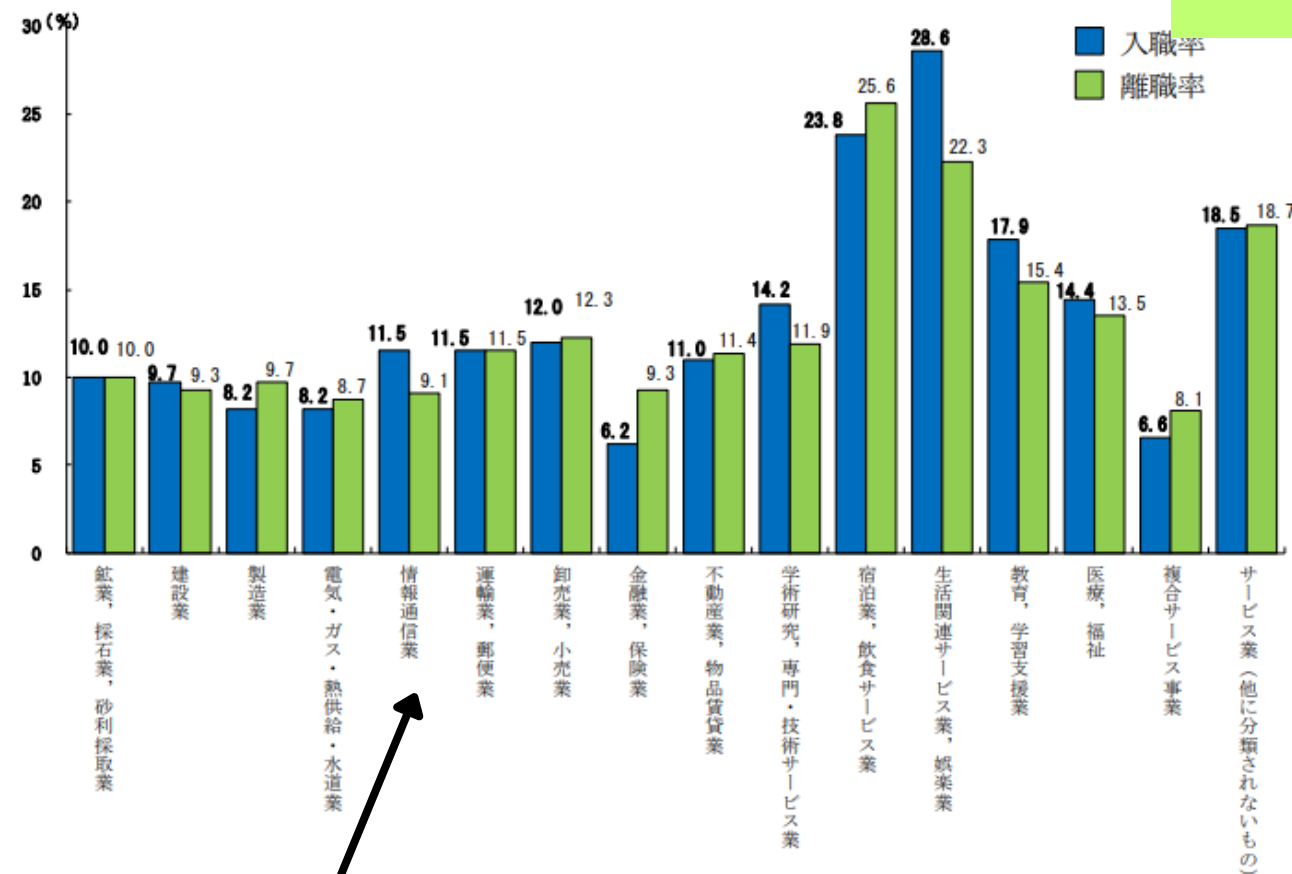


1. 市場分析
2. 問題設定
3. EDA・特徴量エンジニアリング
4. 事業提案

市場分析

情報通信業界の離職率

図3-1 産業別入職率・離職率（令和3年（2021））



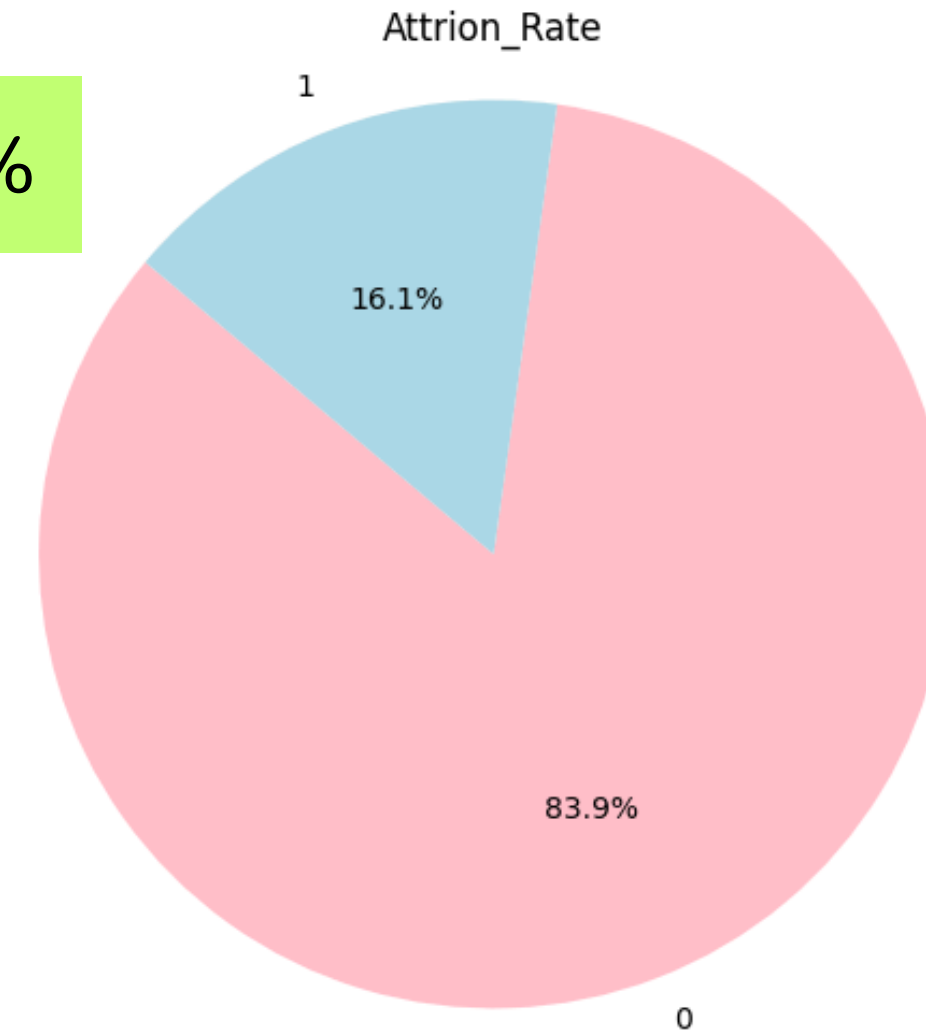
上図：厚生労働省の念雇用動向調査からの引用

情報通信業界では平均9.1%の離職率
他の業界と比較しても低い水準

9.1%

16.1%

7.0%



上図：I社における離職率

情報通信業界の平均と比較してI社は離職率が7.0%高い
（0 が在籍・1 が離職）

問題設定

IT人材の需給推移と予測

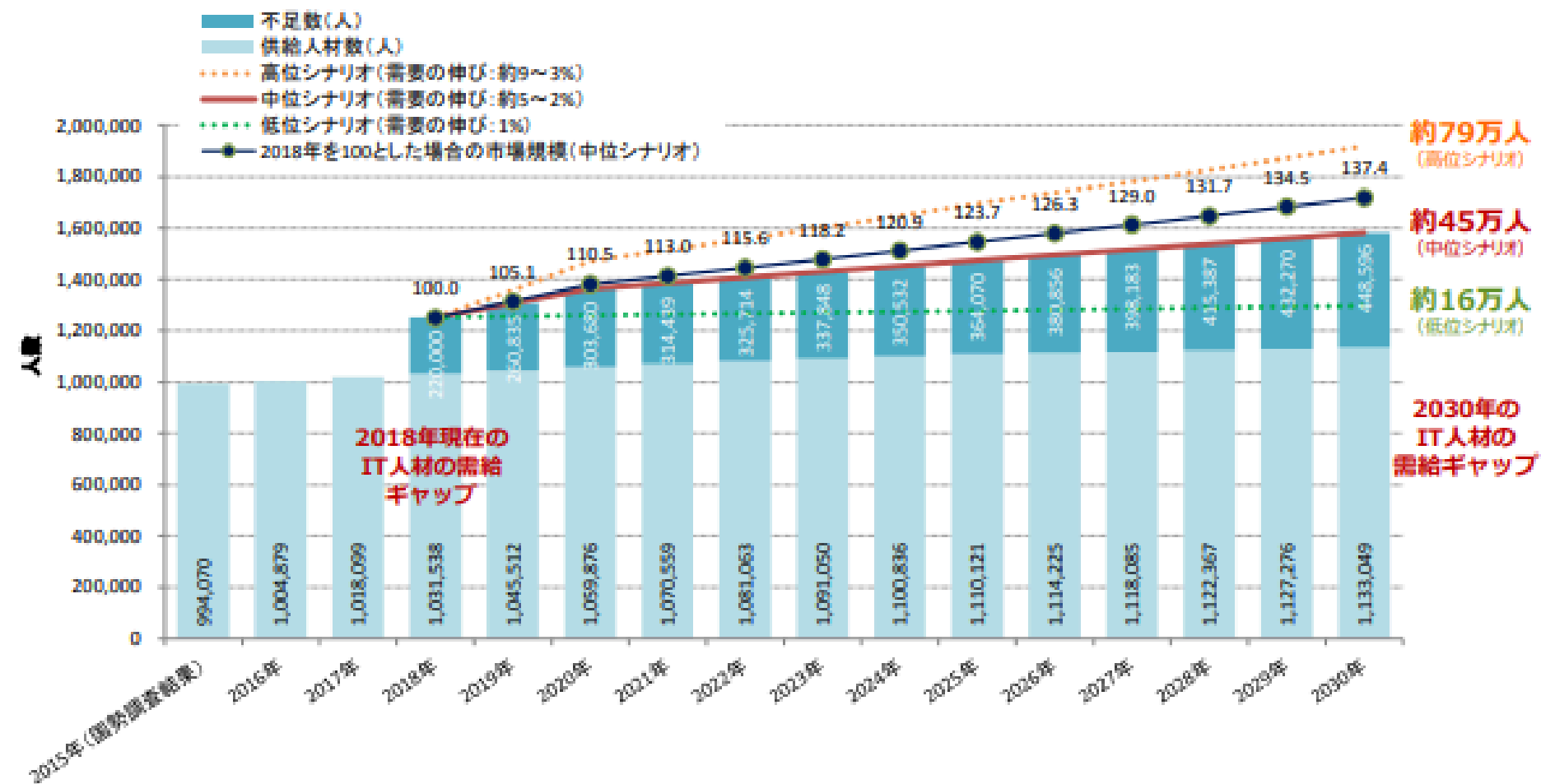


図 3-11 IT 人材需給に関する主な試算結果①②③の対比

(生産性上昇率 0.7%、IT 需要の伸び「低位」「中位」「高位」)

(出所) 2015 年は総務省「平成 27 年国勢調査」によるもの、
2016 年以降は試算結果をもとにみずほ情報総研作成

左図：経済産業省 IT 人材需給に関する調査

需要は高まるが供給が追いつかずに、2030年には最低でも16万人の人材が不足してくることがあげられている。

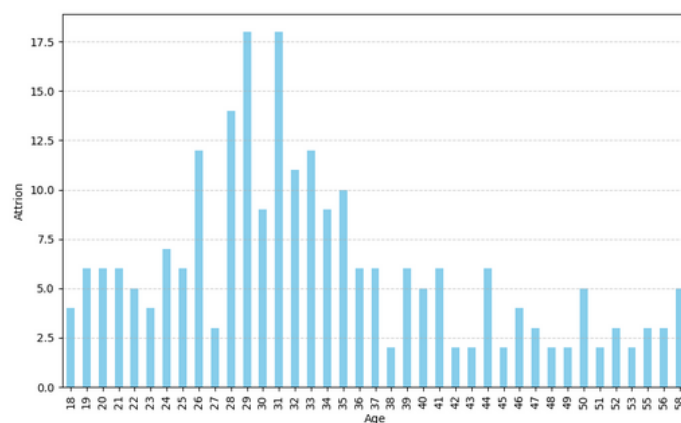
人材不足が見込まれる中で会社内の離職率が高いことは、企業の持続可能性や競争力に深刻な影響を与える可能性があります。

データの活用で
この課題を解決に導くご提案

EDA・特徴量エンジニアリング

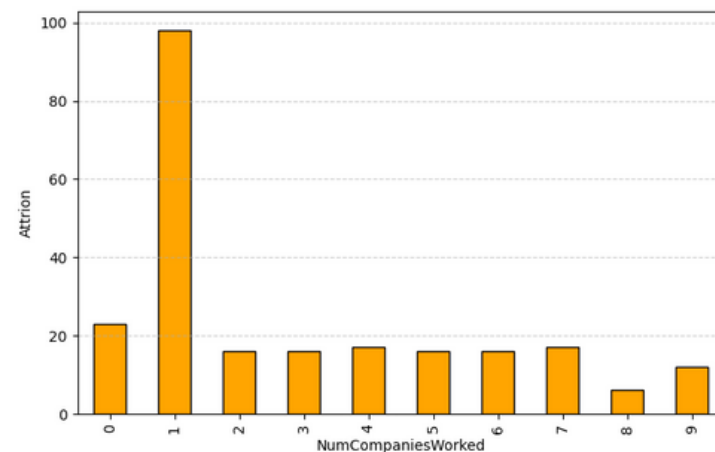
各特徴量と離職の相関

離職と関係の深い特徴量を比較する



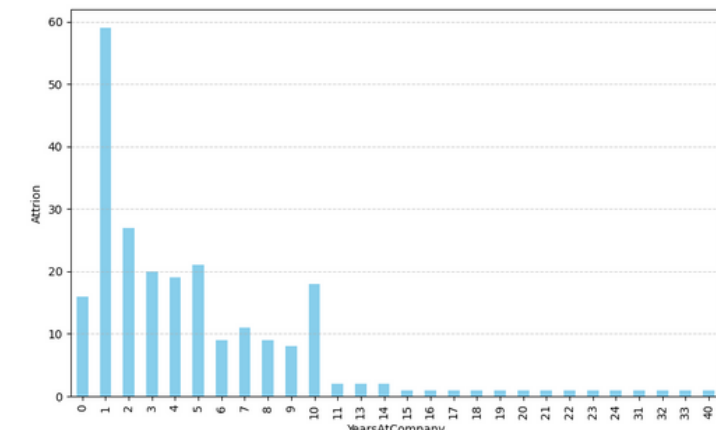
年齢vs離職

20代後半か30代前半が離職者が多い



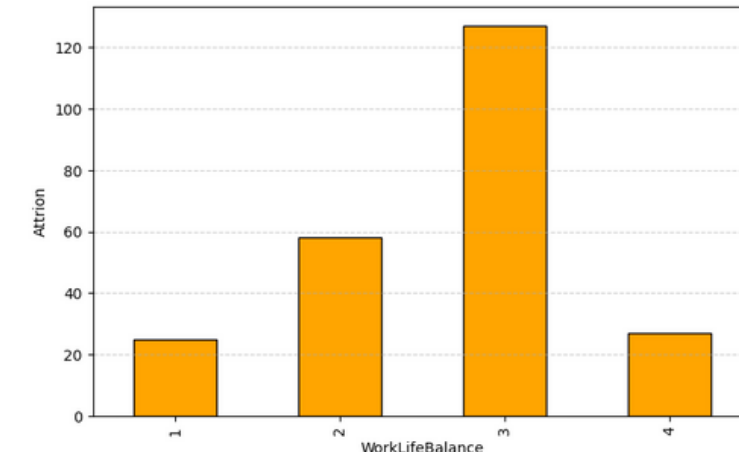
転職回数vs離職

転職が1回の社員に離職者が多い



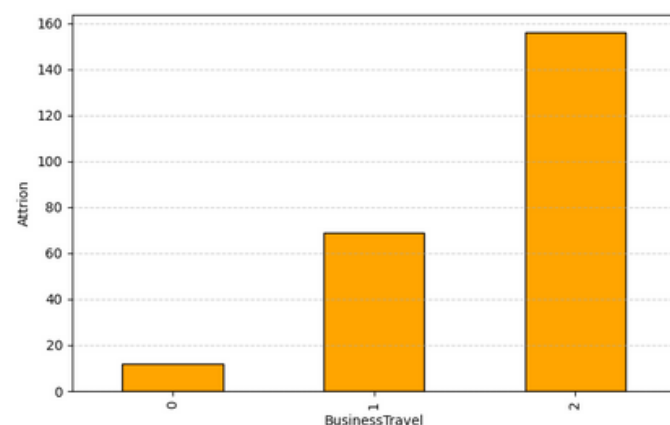
勤続年数vs離職

勤続年数1年で離職する社員が多い



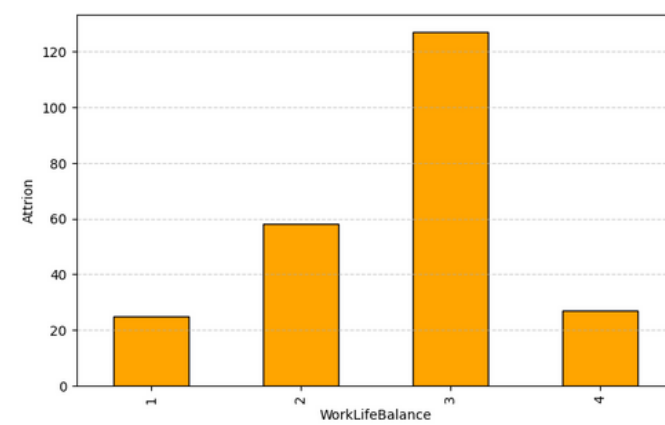
月収vs離職

0k~3kの月収の社員が離職しやすい



出張頻度vs離職

出張が多いと離職が多くなる



ワークライフバランスvs離職

ワークライフバランスが良いと考えている社員に離職者が多い

その他の特徴量でも離職との相関を調査

- DistanceFromHome
 - 1,2,9 自宅から勤務地の距離が短い社員に離職が多い
- Education
 - 3,4,2,1 修士卒の社員に離職者が多い
- EnvironmentSatisfuction
 - 1 環境満足度が悪いと離職者が多い
- JobSatisfaction
 - 3 仕事満足度が良いと離職者が多い
- OverTime
 - 1 残業があると離職者が多い
- Gender
 - 1 女性には離職者が多い

事業提案

離職者予測モデルの提案

ロジスティック回帰

ロジスティック回帰は、「この条件ならこうなる」といったシンプルなルールで予測します。

例えば、「価格が高いと売れにくい」など、直感的でわかりやすい結論を出します。

特徴: 簡単で速く結果が出るが、複雑な問題には弱い。

XGブースト

XGブーストは、一人一人が部分的に予測を改善し、最終的にチーム全体で精度を高めます。

例えば、「価格、競合、季節、地域の全てを考慮して結論を出す」ような複雑な問題に強いです。

特徴: 複雑な問題で高精度だが、計算に時間がかかり結果の説明が難しい。

モデル評価結果

指標	訓練データ	テストデータ
正解率 (Accuracy)	0.802	0.758
精度/適合率 (Precision)	0.794	0.791
再現率 (Recall)	0.815	0.708
F値 (F1-score)	0.804	0.747

テストデータ分類レポート

クラス	精度/適合率 (Precision)	再現率 (Recall)	F値 (F1-score)	サンプル数 (Support)
クラス 0	0.69	0.77	0.73	47
クラス 1	0.74	0.67	0.70	48
平均	0.72	0.72	0.72	95

参考元：<https://qiita.com/dsninaru1250/items/2ba7859154e2b33c0b83>

事業提案

モデルの評価

モデルA

ロジスティック回帰

モデル評価結果

指標	訓練データ	テストデータ
正解率 (Accuracy)	0.802	0.758
精度/適合率 (Precision)	0.794	0.791
再現率 (Recall)	0.815	0.708
F値 (F1-score)	0.804	0.747

モデルB

XGブースト

テストデータ分類レポート

クラス	精度/適合率 (Precision)	再現率 (Recall)	F値 (F1-score)	サンプル数 (Support)
クラス 0	0.69	0.77	0.73	47
クラス 1	0.74	0.67	0.70	48
平均	0.72	0.72	0.72	95

評価の基準

評価指標	目安のスコア	理由
再現率 (Recall)	$\geq 70\%$	離職者を見逃す割合を30%以下に抑える必要があるため。低いとモデルの価値が低下。
F1スコア	≥ 0.7	精度と再現率のバランスが良いと評価できる。
正解率 (Accuracy)	$\geq 75\%$	全体として信頼できるモデルと感じてもらうため。ただし、クラス不均衡には注意が必要。
精度 (Precision)	$\geq 75\%$	離職者として予測した人が実際に離職する割合を高くすることで、無駄なコストを防ぐ。

ロジスティック回帰では、**離職者は少し見逃しやすいが、精度は高いため効率は良い**

XGブーストでは、精度、正解率、F1スコア、再現率の**数値が良く、バランスが良い**

参考元：<https://qiita.com/dsninaru1250/items/2ba7859154e2b33c0b83>

事業提案

汎用性について

また、ロジスティック回帰のハイパーパラメータチューニングの結果は以下の通りです

ベストスコア: **0.7573684210526315**

ハイパーパラメータチューニングとは、モデルが**最も良い性能を発揮できる設定を探す作業**です。

- ハイパーパラメータ: **モデルの動きを調整する設定項目**です（例: 木の深さや学習率）。
 - 目的: ハイパーパラメータを調整することで、モデルが学習データにも**未知のデータにも正確に対応できるようにすること**。
- 具体的には、さまざまなハイパーパラメータの組み合わせを試し、その中から最も良い結果（スコア）を出す組み合わせを選びます。

つまり、ロジスティック回帰の汎化性能は低くないという指標が得られました

今回は
「精度重視の**モデルA**」 と 「バランス型の**モデルB**」
をご提案します



事業提案

事業がもたらす利益

1. 離職リスクを採用の段階から

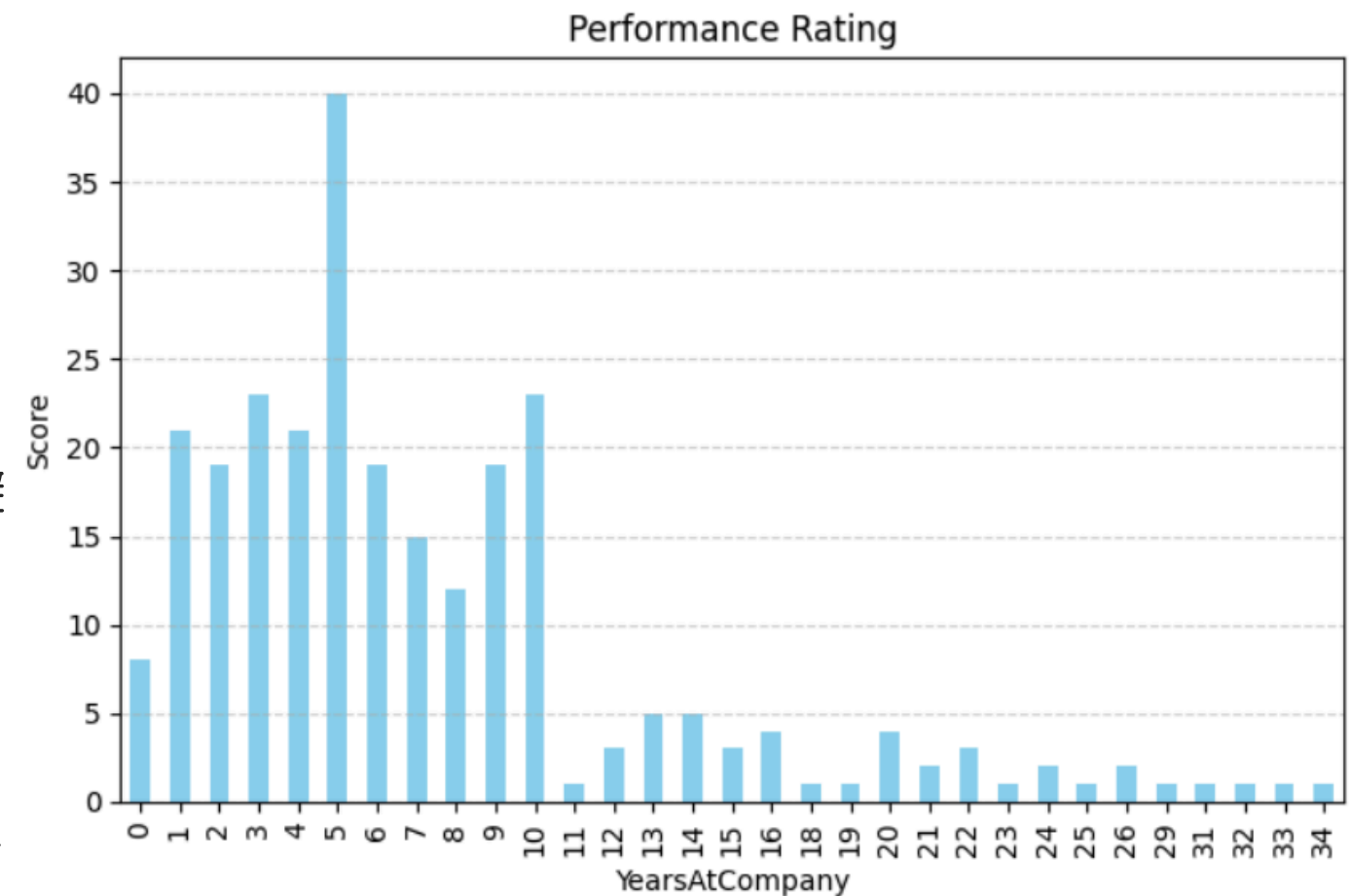
- 採用コストの削減: 離職リスクが高い社員を事前に把握し、適切なフォローを行うことで、不要な離職を防ぎます。
- トレーニングコストの削減: 新人研修などにかかるコストを抑えることができます。

2. 生産性向上が期待できます

- ノウハウの維持: 熟練社員の離職を防ぐことで、プロジェクトの停滞や品質低下を回避します。
 - 御社で業績を上げている社員は1~10年金属している社員に多いです（左図）

3. 離職を適切に防ぐことが出来ます

- 社員満足度の向上: 離職リスクのある社員の要因を解消し、働きやすい環境を整備します。
- イメージアップ: 離職率が低い会社は、外部からの信頼が向上し、優秀な人材を引きつける力が強まります



事業提案

想定されるシナリオ

実際の活用シナリオ

ケース1: 若手社員の早期離職防止

- モデルでリスクが高い新入社員を特定。
- 満足度に離職と関連する値が見られない
- 勤続年数が浅いことに因果を推測
- メンター制度やキャリアパスの再設計を個人用カスタムで導入。

ケース2: 働きすぎによる離職リスクの軽減

- 残業時間が多い社員のデータを分析。
- 環境満足度が悪い（1）だった
- 自動化ツールの導入や人員配置の見直しを実施。



最後に

御社のますますのご発展を祈念しております。
ぜひご検討ください。

ご清聴ありがとうございました。

