ータを売上額で除算する場合で散布図を比較したところ、各データを売上額で除算する場合は散布図の形状が大きく異なることを確認した。このことは、業務知識に基づく前処理が予測モデルの精度に大きな影響を与える可能性があることを示唆している。そのため、予測モデルを作成する際には、エンジニアのみならず業務知識豊富な担当者の参画が重要となる。

ロ　予測モデル作成

損益計算書データのうちランダムに選択した半数に対して売上額を10%削減し、疑似的に誤りのある損益計算書を作成した。同様の方法で、売上額を 5%、2.5%削減したデータセット及び仕入額を 10%、5%、2.5%増加させたデータセットの合計６種類のデータセットを作成した。当該データセットに対して、６種類のスケーリング手法及び５種類の機械学習手法（ロジスティック回帰、サポートベクターマシン、決定木、ランダムフォレスト及び深層学習）を組み合わせて予測モデルを構築し、その予測精度を比較した。

６種類のデータセットに対する予測精度の最高値の達成状況は、サポートベクターマシン２回、ランダムフォレスト２回、決定木１回であった。しかしながら、６種類のデータセットと６種類のスケーリング方法の組み合わせである 36 のケースにおける予測精度について、機械学習ごとに平均値を算出すると、深層学習以外は全て 68%台であり、大きな差は認められなかった。このことは、サポートベクターマシン及びランダムフォレストは高精度を示すことが多いものの、安定性を欠いたことを示唆している。同一の損益計算書から作成したデータセットであっても、売上額や仕入額の削減・増加割合の違いとスケーリング手法の違いによって最も高い予測精度を達成するアルゴリズムが異なったことが示すように、最適な機械学習手法を事前に選択することは困難である。そのため、実務において予測モデルを作成する際には、利用可能な機器の性能、求められる説明性及び解釈性を勘案した上で、複数のモデルを