

福岡工業大学 令和7年度 卒業研究論文

授業評価の数値に表れない学生の本音  
—マルチタスク学習と SHAP 分析による満足度要因の解明—

指導教員 佐藤 大輔

福岡工業大学情報工学部  
システムマネジメント学科

学籍番号 22M11178

氏名 藺牟田 晃弘



# 目次

第1章 はじめに	1
1.1 研究背景	1
1.2 課題の整理	1
1.3 研究目的	1
1.4 研究仮説	2
1.5 研究のアプローチと特徴	2
1.6 研究の意義	2
1.7 本研究の構成	2
第2章 関連研究	4
2.1 授業評価研究	4
2.2 自由記述分析と感情分析	4
2.3 自由記述分析の手法分類	5
2.4 BERT と日本語感情分析	5
2.5 マルチタスク学習	5
2.6 解釈可能 AI と SHAP	6
2.7 順序回帰	6
2.8 教育分野への応用研究の整理	6
2.9 既存研究の限界と課題	7
2.10 本研究の位置づけ	7
第3章 データと手法	8
3.1 データセット	8
3.2 授業評価アンケートの構成	8

3.3	前処理	8
3.4	教師データ	9
3.5	モデル構成	9
3.5.1	感情分類モデル	9
3.5.2	マルチタスク学習モデル	9
3.5.3	順序回帰モデル	10
3.6	学習設定	10
3.7	授業単位集約と相関分析	10
3.8	SHAP 分析	11
3.9	評価指標	11
3.10	分析フロー	11
<b>第 4 章</b>	<b>結果と考察</b>	<b>12</b>
4.1	基礎統計量	12
4.2	感情スコアと授業評価スコアの相関分析	12
4.3	感情分類モデルの性能	12
4.4	単一タスクモデルの SHAP 分析	12
4.5	マルチタスク学習の SHAP 分析	13
4.6	感情特化要因・評価特化要因の示唆	15
4.7	順序回帰モデルの結果（追加実験）	16
4.8	総合考察	16
<b>第 5 章</b>	<b>おわりに</b>	<b>19</b>
5.1	結論	19
5.2	実践的示唆	19
5.3	限界	19
5.4	今後の課題	20
<b>付 録 A</b>	<b>感情ラベル定義と例</b>	<b>21</b>
<b>付 録 B</b>	<b>データセット詳細</b>	<b>23</b>



# 図 目 次

3.1 分析フローの概略 . . . . .	11
4.1 要因タイプの割合（概念図） . . . . .	14

# 表 目 次

3.1	データセット概要 . . . . .	9
3.2	教師データのラベル分布 (1,000 件) . . . . .	10
4.1	感情スコアと授業評価スコアの基本統計量 . . . . .	13
4.2	教師データのラベル分布 (1,000 件) . . . . .	14
4.3	感情スコアと授業評価スコアの相関分析結果 . . . . .	15
4.4	感情分類モデルの性能指標 . . . . .	16
4.5	ポジティブ判定に寄与する重要語 TOP10 . . . . .	17
4.6	語彙の要因タイプ別内訳 . . . . .	17
4.7	要因タイプの解釈例 . . . . .	18
4.8	共通要因 (満足度要因) TOP5 . . . . .	18
A.1	感情ラベルの定義 . . . . .	22
B.1	データセット概要 . . . . .	24
B.2	教師データのラベル分布 (1,000 件) . . . . .	24

# 第1章 はじめに

## 1.1 研究背景

高等教育機関では、教育の質向上を目的として学生による授業評価が広く実施されている。授業評価アンケートは、多段階の評価スコアと自由記述から構成されることが多く、評価スコアは定量的な比較に適している。一方で、授業の良否を左右する要因や学生の本音は自由記述に表れることが多く、数値評価のみでは十分に把握できない。

近年、自然言語処理技術と機械学習の発展により、大規模な自由記述の分析が可能になってきた。自由記述を感情スコアとして数値化し、授業評価スコアとの関係を統計的に検討することで、授業改善に有用な知見を得られる可能性がある。

## 1.2 課題の整理

授業評価の活用には、以下の課題が存在する。第一に、授業評価スコアは複数要因の総合判断であり、どの要因がどの程度影響したかが直接観測できない点である。第二に、自由記述は非構造データであるため、人的読解に依存すると大規模データの分析が困難である。第三に、授業評価スコアのみでは学生の感情や満足度の差異を捉えにくく、教育改善に結び付く具体的示唆が得られにくい点が挙げられる。

## 1.3 研究目的

本研究の目的は、授業評価アンケートの自由記述から感情スコアを推定し、授業評価スコアとの関係性を分析することで、授業評価に影響する要因を定量的に特定することである。

具体的には、(1) 自由記述の感情分析により感情スコアを算出し、授業単位で集計した感情スコアと授業評価スコアの関係を検討する、(2) 感情スコアと評価スコアを同時に予測するモデルを構



築し、共通要因（満足度要因）とそれぞれに特有の要因を分離する、(3) 解釈可能性手法を用いて要因を定量化し、授業改善に資する知見を得る、という流れで目的の達成を図る。

## 1.4 研究仮説

本研究では、以下の仮説を設定する。

1. 授業単位で集約した感情スコアと授業評価スコアには正の相関関係がある。
2. 感情スコアと授業評価スコアの両方に影響する共通要因（満足度要因）が存在する。
3. マルチタスク学習により、共通要因と特化要因を分離できる。

## 1.5 研究のアプローチと特徴

本研究では、日本語の事前学習済み BERT モデルを基盤とした感情分類モデルを構築し、授業評価アンケートの自由記述から感情スコアを推定する。推定した感情スコアは授業単位で集計し、授業評価スコアとの相関関係を統計的に検討する。

さらに、感情スコアと授業評価スコアを同時に予測するマルチタスク学習モデルを構築し、両者に共通する要因と特化要因を分離する。加えて、SHAP 分析を用いて単語レベルの寄与度を定量化し、授業改善に直結する要因の抽出を行う。

## 1.6 研究の意義

本研究は、自由記述を感情スコアとして定量化し、授業評価スコアとの関係を明確化する点に意義がある。また、マルチタスク学習と SHAP 分析を組み合わせることで、単なる相関の確認にとどまらず、授業改善に資する具体的な要因を抽出できる。これにより、教育改善の意思決定をデータに基づいて行うための基盤を提供する。

## 1.7 本研究の構成

本研究は、全 5 章からなる。第 2 章では、授業評価研究、自然言語処理による感情分析、マルチタスク学習、解釈可能 AI、順序回帰に関する関連研究を整理する。第 3 章では、データセット

と前処理，感情分類モデル，マルチタスク学習，順序回帰モデル，および評価方法を述べる．第4章では，感情スコアと授業評価スコアの関係性分析，モデルの推定結果，SHAP 分析による要因抽出の結果を示す．第5章では，本研究のまとめと今後の課題を述べる．

## 第2章 関連研究

### 2.1 授業評価研究

大学における授業評価は、教育の質向上に向けた重要な指標として用いられている。多くの大学では、学期末のアンケートにより授業評価スコアと自由記述を収集し、教員へフィードバックを行っている。授業評価スコアは数量的に扱いやすい一方で、学生が評価に至った理由や具体的な改善要望は自由記述に含まれることが多い。このため、自由記述を定量的に分析し、授業評価スコアの背後にある要因を明らかにする研究が必要とされている

citehujala2020,santhanam2018. さらに、授業改善への活用を前提とした分析プロセスの標準化や、フィードバックの迅速化が課題として指摘されており、効率的な分析手法の整備が求められる citehujala2020.

### 2.2 自由記述分析と感情分析

自由記述は非構造テキストであり、従来は人的な読解に依存していた。しかし、近年の自然言語処理技術の発展により、大規模な自由記述を自動的に解析し、感情や評価の傾向を抽出することが可能になっている。感情分析は、テキストに含まれる肯定的・否定的・中立的な感情を推定する技術であり、教育分野においても学生の満足度や不満の把握に活用できる

citeraajput2016,sindhu2019. 教育分野の自由記述分析では、テキスト分析を通じた改善提案の抽出や意見整理が報告されており

citegottipati2018,misuraca2021, 従来のスコア中心の評価を補完する手段としての重要性が高まっている。

## 2.3 自由記述分析の手法分類

自由記述分析は大きく、(1) 辞書型手法、(2) 古典的機械学習手法、(3) 深層学習手法に分類できる。辞書型手法は、極性語彙を用いて感情を推定するため実装が容易であるが、文脈依存の表現や否定表現に弱い

citeraajput2016. 古典的機械学習手法は、特徴量設計により一定の精度を得られるが、語彙の多様性が高い自由記述では特徴量設計の負担が大きい

citesanthanam2018. 深層学習手法は文脈を考慮できる一方で、教師データの準備コストが高く、教育分野固有の語彙や表現への適応が課題となる

citebert.

## 2.4 BERT と日本語感情分析

BERT (Bidirectional Encoder Representations from Transformers) は、Transformer を基盤とした事前学習済み言語モデルである

citetransformer. 双方向の文脈情報を利用できる点が特長であり、少量の教師データでも高精度な分類が可能であると報告されている

citebert. 日本語に対しても事前学習済みモデルが提供されており、教育分野の自由記述に対して微調整を行うことで、文脈を考慮した感情分類が実現できる。一方で、学習データの領域差が大きい場合には汎化性能が低下する可能性があるため、教育分野に特化した微調整と評価設計が必要となる。

## 2.5 マルチタスク学習

マルチタスク学習は、複数の関連するタスクを同時に学習し、共通表現を獲得することで各タスクの性能を向上させる手法である

citemtl. 感情分析と授業評価スコア予測は、いずれも自由記述に基づく評価理解という共通の目的を持つため、マルチタスク学習の適用が有効であると考えられる。特に、感情スコアと評価スコアを同時に学習することで、共通要因と特化要因の分離が可能となり、教育改善の示唆をより具体化できる点が期待される。

## 2.6 解釈可能 AI と SHAP

機械学習モデルの予測根拠を明確化するため、解釈可能 AI (XAI) が注目されている。SHAP (SHapley Additive exPlanations) は、協力ゲーム理論に基づき、特徴量の寄与度を定量化する手法である

citeshap. テキスト分類においては、単語レベルの寄与度を算出できるため、授業評価に影響する要因を具体的な語彙として提示できる。教育分野では、モデルの予測精度だけでなく、説明可能性が重要であり、改善施策への翻訳可能性が求められる。SHAP はこの点で有用であるが、寄与度の解釈は文脈に依存するため、定性的な検討との併用が必要となる

citeshap.

## 2.7 順序回帰

授業評価スコアは 1 点から 4 点までの順序尺度であり、単純な回帰や分類では順序関係を適切に扱えない。順序回帰は、評価段階の順序性を考慮して確率分布を推定する手法であり、近年はニューラルネットワークと組み合わせた手法も提案されている

citecoral. 授業評価スコアの分析においても、順序回帰の導入は妥当である。

## 2.8 教育分野への応用研究の整理

教育分野では、学習ログやアンケートデータを用いた分析が進んでいる。自由記述を対象としたテキスト分析や意見抽出の取り組みは報告されているが

citegottipati2018,misuraca2021,hujala2020, 評価スコアとの関係性を統合的に扱った研究は多くない。また、教育改善に直結する語彙や要因を整理するために、語彙辞書の拡張や分析手法の整理が試みられている

citesanthanam2018. このため、感情分析・マルチタスク学習・解釈可能 AI を組み合わせた総合的な分析枠組みの構築が求められている。

## 2.9 既存研究の限界と課題

既存研究は、(1) 評価スコアと自由記述の統合が不十分、(2) 感情分析結果の解釈が定性的で、教育改善に直結しづらい、(3) 予測精度と説明可能性の両立が難しい、という課題を持つ。特に、評価スコアと感情情報の関係を同時にモデル化した研究は限られており、改善施策を定量的に導出するための枠組みが不足している

citegottipati2018,misuraca2021.

## 2.10 本研究の位置づけ

本研究は、授業評価アンケートの自由記述に対し、BERT による感情分類とマルチタスク学習を適用し、さらに SHAP 分析によって評価要因を定量化する点に特徴がある。特に、感情スコアと授業評価スコアを同時に学習することで、共通要因と特化要因を分離し、改善施策の優先順位付けに利用できることを目指す。これにより、教育改善に資する具体的な知見を提供することを目指す。

## 第3章 データと手法

### 3.1 データセット

本研究では、福岡工業大学の授業評価システムにおける 2018 年度から 2023 年度までのデータを使用した。対象は 9 学科、授業数は 3,268 件である。自由記述の総件数は 83,851 件であり、各授業に対して複数の自由記述が付随する。

データセットの概要を表 3.1 に示す。

### 3.2 授業評価アンケートの構成

授業評価アンケートは、(1) 択一式質問の点数化による授業評価スコア、(2) 自由記述の 2 種類の情報から構成される。自由記述は以下の 2 つの質問からなる。

1. 先生に向けてこの授業の感想や学んだこと、意見や要望を記述してください
2. 次期履修者に向けて、この授業についてのアドバイスを記述してください

授業評価スコアは 4 段階（1～4 点）であり、授業ごとに単一のスコアが付与される。一方、自由記述は授業単位で複数件存在するため、授業単位で集約して分析する必要がある。

### 3.3 前処理

自由記述は、日本語形態素解析によりトークン化を行い、記号や不要な空白の除去などの正規化を施した。その後、モデル入力に適した形式へ変換し、学習データとして利用した。

表 3.1: データセット概要

項目	値
対象期間	2018 年度～2023 年度
対象学科数	9
授業数	3,268
自由記述総件数	83,851
平均自由記述数/授業	25.2

### 3.4 教師データ

感情分類モデルの構築のため、ランダム抽出した 1,000 件の自由記述を手動でラベリングし、教師データとして用いた。ラベルはネガティブ、ポジティブ、ニュートラルの 3 クラスとし、分類結果はネガティブ 191 件、ポジティブ 180 件、ニュートラル 628 件であった。

教師データのラベル分布を表 3.2 に示す。

### 3.5 モデル構成

#### 3.5.1 感情分類モデル

感情分類には、日本語の事前学習済み BERT モデルを用いた。BERT の出力表現に分類器を接続し、3 クラス分類として微調整を行った。これにより、自由記述ごとの感情スコア（ネガティブ=-1, ニュートラル=0, ポジティブ=+1）を推定した。

#### 3.5.2 マルチタスク学習モデル

感情スコア予測と授業評価スコア予測を同時に学習するマルチタスクモデルを構築した。BERT エンコーダを共有し、感情分類ヘッドと評価スコア予測ヘッドを分岐させる構成とした。これに



表 3.2: 教師データのラベル分布 (1,000 件)

ラベル	件数
ネガティブ	191
ニュートラル	628
ポジティブ	180

より，共通要因と特化要因の分離が可能となる．

### 3.5.3 順序回帰モデル

授業評価スコアは順序尺度であるため，順序回帰モデルを構築する．順序回帰モデルでは，評価スコア 1～4 点の確率分布 (P1, P2, P3, P4) を同時に推定し，評価段階ごとの特徴を分析する．本節の詳細は，追加実験の完了後に追記する．

## 3.6 学習設定

モデル学習は，教師データを訓練用・検証用に分割して実施した．学習率やバッチサイズは予備実験により調整し，過学習を防ぐために早期終了などの設定を用いた．

## 3.7 授業単位集約と相関分析

自由記述ごとの感情スコアを授業単位で平均し，授業単位の感情スコアを算出した．この感情スコアと授業評価スコアの関係性を検討するため，ピアソン相関係数，スピアマン順位相関係数，

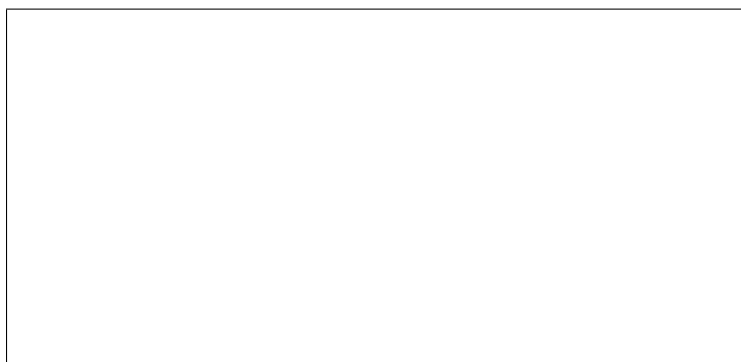


図 3.1: 分析フローの概略

ケンドール順位相関係数を用いて相関分析を行った。

### 3.8 SHAP 分析

モデルの解釈可能性を高めるため、SHAP 分析を実施した。単一タスクモデルおよびマルチタスクモデルに対し、層化サンプリングで 5,000 件を抽出し、出現回数 5 回以上の単語を対象として寄与度を算出した。最終的に 1,564 語を対象に分析を行った。

### 3.9 評価指標

感情分類の性能評価には正解率および F1 スコアを用いた。授業評価スコア予測の性能評価には決定係数 ( $R^2$ ) を用いた。相関分析では有意確率を併記し、統計的有意性を確認した。

### 3.10 分析フロー

分析全体の流れを図 3.1 に示す。

## 第4章 結果と考察

本章では、授業評価アンケートの基礎統計量、相関分析結果、SHAP 分析による要因抽出結果を示し、得られた知見を考察する。

### 4.1 基礎統計量

感情スコアと授業評価スコアの基本統計量を表 4.1 に示す。感情スコアは平均 0.001 でほぼニュートラルに近く、授業評価スコアは平均 3.459 点（4 点満点）で比較的高い評価が多いことが分かる。

教師データのラベル分布を表 4.2 に示す。ニュートラルが大半を占める一方、ネガティブ・ポジティブは少数であり、クラス不均衡が存在する。

### 4.2 感情スコアと授業評価スコアの相関分析

授業単位で集約した感情スコアと授業評価スコアの相関分析結果を表 4.3 に示す。ピアソン相関係数は 0.3097 ( $p < 0.000001$ ) であり、中程度の正の相関が確認された。順位相関係数でも同様に正の相関が得られ、複数の指標で一貫した結果となった。

### 4.3 感情分類モデルの性能

感情分類モデルの性能指標を表 4.4 に示す。詳細な数値は最終的な学習条件の確定後に追記する。

### 4.4 単一タスクモデルの SHAP 分析

感情分類モデルに対する SHAP 分析を行い、ポジティブ判定に寄与する重要語を抽出した。上位 10 語を表 4.5 に示す。「やす」「良かつ」「おもしろ」「面白」など、理解のしやすさや授業の面白さを示す語彙が上位に現れた。

表 4.1: 感情スコアと授業評価スコアの基本統計量

統計量	感情スコア	授業評価スコア
平均	0.001	3.459
標準偏差	0.260	0.216
最小値	-1.000	2.000
第 1 四分位数 (Q1)	-0.167	3.330
中央値 (Q2)	0.000	3.480
第 3 四分位数 (Q3)	0.167	3.600
最大値	1.000	4.000

理解のしやすさや面白さに関連する語彙が上位に位置することから、学生の満足度には内容理解と興味喚起が大きく関与することが示唆される。特に「やす」は「分かりやすい」「理解しやすい」などに共通する語幹であり、授業理解のしやすさが肯定的評価の中心的要因である可能性を示す。また、「面白」「おもしろ」「楽しい」といった語彙が複数ランクインしている点は、授業内容の興味・関心が感情評価に与える影響が大きいことを示唆する。これらの結果は、評価スコアが高い授業ほど「理解しやすさ」と「面白さ」が同時に高まるという仮説と整合的であり、授業改善においては内容構成の明確化や関心を引く説明の設計が有効な方向性となる。一方で、単一タスクの SHAP 分析は感情スコアに対する寄与を示すものであり、評価スコアへの直接的な影響を意味するものではない点に留意が必要である。

## 4.5 マルチタスク学習の SHAP 分析

マルチタスクモデルの SHAP 分析により、感情スコアと授業評価スコアの両方に影響する共通要因と、各タスクに特化した要因を分離した。語彙の分類結果を表 4.6 に示す。共通要因は 577 語 (18.0%) であり、満足度要因として授業改善への示唆が得られる。

共通要因の上位語彙を表 4.8 に示す。「学ぶ」「理解」「総括」「推奨」など、学習成果に関わる語彙が上位に現れている。

共通要因は、感情スコアと評価スコアの双方に同時に寄与する語彙であり、限られた資源で授

表 4.2: 教師データのラベル分布 (1,000 件)

ラベル	件数
ネガティブ	191
ニュートラル	628
ポジティブ	180



図 4.1: 要因タイプの割合 (概念図)

業改善を行う際の「投資効率が高い要因」と解釈できる。例えば「学ぶ」「理解」「総括」といった語彙は、授業内容の理解度や学習成果の実感を示唆するため、学生にとって「学びが得られた」という感覚が満足度と評価の双方に影響している可能性が高い。感情特化要因には「楽しい」「面白い」など情緒的評価を反映する語彙が多く含まれると考えられ、これらは満足感の形成には強く寄与するが、評価スコアへの影響は必ずしも大きくない可能性がある。一方で、評価特化要因は授業設計や運営の実務的な要素に関わる語彙が含まれる可能性が高く、評価スコアの改善には直接的に寄与するが、感情面の改善には限定的となることが示唆される。

このように、マルチタスク学習によって「共通要因」「感情特化要因」「評価特化要因」を分離できる点は、授業改善の方針を選択する際に有用である。具体的には、満足度と評価の両方を高

表 4.3: 感情スコアと授業評価スコアの相関分析結果

指標	相関係数	$p$ 値
ピアソン相関係数	0.3097	$< 0.000001$
スピアマン順位相関係数	0.2970	$< 0.000001$
ケンドール順位相関係数	0.2042	$< 0.000001$

めたい場合は共通要因に対する改善を優先し、感情的満足を高めたい場合は感情特化要因を、評価指標を改善したい場合は評価特化要因を重点的に改善する戦略が考えられる。

さらに、共通要因の割合が 18.0%にとどまる点は、満足度と評価が必ずしも完全に一致するわけではないことを示している。これは、評価スコアが授業の「成果」や「有用性」を反映しやすい一方で、感情スコアは「楽しさ」や「雰囲気」といった情緒的側面を反映しやすいことによるものと解釈できる。そのため、評価スコアのみでは把握しきれない学生の感情的側面を補う指標として、感情スコアを活用する意義がある。

以上より、マルチタスク学習と SHAP 分析の組み合わせは、授業改善における具体的な指針を提供するだけでなく、評価スコアと感情スコアの関係性をより精緻に理解するための有効な枠組みであるといえる。

## 4.6 感情特化要因・評価特化要因の示唆

感情特化要因は、学生の感情的評価を強く反映する語彙群であり、授業の雰囲気や楽しさ、満足感に関わる表現が中心となる傾向がある。これらは感情スコアの上昇に寄与するが、評価スコアへの影響は必ずしも大きくない可能性がある。

一方、評価特化要因は授業の有用性や学習成果、授業運営の評価に関わる語彙が中心となると考えられる。これらは評価スコアの改善に直結しやすいが、感情面の改善には限定的となる可能性がある。

この区別により、「満足感を高める施策」と「評価スコアを高める施策」を分けて設計できる点が、本研究の結果の実践的意義である。

表 4.4: 感情分類モデルの性能指標

指標	値
正解率	—
F1 スコア	—

#### 4.7 順序回帰モデルの結果（追加実験）

順序回帰モデルの結果は、追加実験として現在検証中である。本節では、評価段階ごとの寄与語や P2（中低評価確率）、P4（高評価確率）の差異を中心に整理する予定である。結果の確定後に表・図を追加する。

#### 4.8 総合考察

相関分析により、授業レベルでは感情スコアと授業評価スコアに統計的に有意な正の相関が確認された。これは、学生の自由記述に表れる感情が授業評価と一定の関係を持つことを示している。

SHAP 分析から、理解しやすさや授業の面白さに関わる語彙がポジティブ判定に強く寄与することが示された。また、共通要因では「学ぶ」「理解」「総括」など学習成果に関わる語彙が上位にあり、感情と評価の双方を高める可能性が示唆された。

一方、本研究の結果は相関関係に基づくものであり、因果関係を主張するものではない。今後は実験的検証や介入研究により、因果的な関係の検証が必要である。

表 4.5: ポジティブ判定に寄与する重要語 TOP10

順位	単語	平均 SHAP 値	出現回数
1	やす	0.2660	337
2	良かつ	0.2466	207
3	おもしろ	0.2438	10
4	よかつ	0.2251	195
5	面白	0.2178	100
6	楽しい	0.1959	67
7	楽しめる	0.1876	6
8	ありが	0.1760	19
9	楽し	0.1642	192
10	面白い	0.1518	37

表 4.6: 語彙の要因タイプ別内訳

要因タイプ	語彙数	割合
共通要因（満足度）	577	18.0%
感情特化要因	1,200	37.5%
評価特化要因	532	16.6%
低重要度要因	889	27.8%



表 4.7: 要因タイプの解釈例

要因タイプ	解釈の方向性	代表語例
共通要因（満足度）	感情と評価の双方に影響	学ぶ, 理解, 総括
感情特化要因	感情の高低を強く反映	楽しい, 面白い
評価特化要因	評価スコアに特に関与	推奨, 人数
低重要度要因	影響が限定的	(該当語彙多数)

表 4.8: 共通要因（満足度要因）TOP5

順位	単語	感情重要度	評価重要度
1	学ぶ	0.001278	0.001386
2	理解	0.001073	0.000833
3	総括	0.000974	0.000952
4	推奨	0.001132	0.000755
5	人数	0.001195	0.000704

## 第5章 おわりに

### 5.1 結論

本研究では、授業評価アンケートの自由記述に対して感情分析を適用し、授業評価スコアとの関係性を分析した。授業単位で集約した感情スコアと授業評価スコアには統計的に有意な正の相関が確認され、学生の感情が授業評価と一定の関係を持つことが示された。

さらに、BERT を用いた感情分類モデルとマルチタスク学習モデルを構築し、SHAP 分析により要因を定量化した。その結果、理解しやすさや面白さに関わる語彙がポジティブ判定に強く寄与することが明らかになった。また、共通要因として「学ぶ」「理解」「総括」「推奨」などが抽出され、感情スコアと授業評価スコアの双方を高める要因であることが示唆された。これらの結果は、授業改善における重点領域の把握に有用である。

### 5.2 実践的示唆

本研究で抽出された共通要因は、教育改善における優先度の高い領域を示す。授業内容の理解を促進する工夫や、興味・関心を高める設計が、満足度向上に寄与する可能性がある。また、評価特化要因や感情特化要因を切り分けることで、授業改善の施策をより精緻に設計できる。

### 5.3 限界

本研究は相関関係の探索を目的としており、因果関係の検証は行っていない。また、データは単一大学に限定されているため、他大学への一般化可能性には限界がある。教師データは1,000件と比較的少なく、ラベル付けの主観性も残る。

## 5.4 今後の課題

今後は、介入研究や実験的検証により因果的関係の解明が必要である。複数大学のデータを用いた比較や、学部・学科ごとの分析を行うことで一般性を高める必要がある。また、教師データの拡充や半教師あり学習の導入により感情分類の精度向上が期待される。さらに、感情特化要因・評価特化要因の活用方法を教育実践に落とし込むための検討が今後の課題である。順序回帰モデルの追加実験を通じて、評価段階ごとの要因差を明らかにすることも重要な課題である。

## 付 録 A 感情ラベル定義と例

本研究で用いた感情ラベルの定義と代表例を表 A.1 に示す．ラベル付けは自由記述の感情的な表現に基づき，人手で分類した．

表 A.1: 感情ラベルの定義

ラベル	感情スコア	定義	例
ネガティブ	-1	不満・批判・否定的な感情を含む記述	この授業はつまらない
ニュートラル	0	事実記述で感情が明確でない記述	板書が多い
ポジティブ	+1	満足・肯定的な感情を含む記述	とても楽しい授業だった

## 付 録 B データセット詳細

本研究で使⽤したデータセットの概要を表 B.1 に示す.

教師データのラベル分布を表 B.2 に示す.

表 B.1: データセット概要

項目	値
対象期間	2018 年度～2023 年度
対象学科数	9
授業数	3,268
自由記述総件数	83,851
平均自由記述数/授業	25.2

表 B.2: 教師データのラベル分布 (1,000 件)

ラベル	件数
ネガティブ	191
ニュートラル	628
ポジティブ	180

# 謝辞

本卒業研究を無事に完了できたのは、多くの方々のご指導とご支援のおかげであり、心より感謝申し上げます。

まず、日々ご指導いただいた高橋先生に深く感謝いたします。研究テーマの選定から論文作成に至るまで、親身にご指導いただきました。特に初期段階では、適切な助言をいただき研究の方向性を定めることができました。また、専門的な助言だけでなく、私の成長を促す課題を与えてくださり、それを乗り越える中で知識やスキルを向上させることができました。先生のご指導なくして本研究の成果は得られなかったと確信しております。

また、研究室の仲間にも感謝いたします。日々の意見交換や協力を通じて、多くの刺激を受け、視野を広げることができました。互いに研究の進捗を共有し、課題を乗り越える中で得た経験は、大学生活における貴重な財産となりました。研究活動以外でも、共に過ごした時間はかけがえない思い出となりました。

さらに、家族の支えなしには研究に専念することはできませんでした。特に両親には、学費や生活面での支援だけでなく、精神的な支えをいただきました。日々の何気ない会話や励ましの言葉が心の支えとなり、困難に直面した際にも家族の存在が大きな励みとなりました。改めて、家族の支えがどれほど大きな力となっていたかを実感しております。

最後に、大学生活を通じて関わったすべての方々に心より感謝申し上げます。先生方や職員の皆様、そして友人たちの支えがあったからこそ、充実した学生生活を送り、多くの学びと経験を得ることができました。これからも学び続け、支えてくださった皆様への感謝を忘れずに精進してまいります。

改めて、この場を借りて深く御礼申し上げます。



## 参考文献

- [1] Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K.: “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pp. 4171–4186 (2019).
- [2] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I.: “Attention Is All You Need,” *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)* (2017).
- [3] Lundberg, S. M., and Lee, S.-I.: “A Unified Approach to Interpreting Model Predictions,” *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)* (2017).
- [4] Zhang, Y., and Yang, Q.: “A Survey on Multi-Task Learning,” *IEEE Transactions on Knowledge and Data Engineering*, Vol. 34, No. 12, pp. 5586–5609 (2022).
- [5] Cao, W., Mirjalili, V., and Raschka, S.: “Rank Consistent Ordinal Regression for Neural Networks with Application to Age Estimation,” *Pattern Recognition Letters*, Vol. 140, pp. 325–331 (2020).
- [6] Gottipati, S., Shankararaman, V., and Lin, J. R.: “Text Analytics Approach to Extract Course Improvement Suggestions from Students’ Feedback,” *Research and Practice in Technology Enhanced Learning*, Vol. 13, Article 6 (2018).
- [7] Rajput, Q., Haider, S., and Ghani, S.: “Lexicon-Based Sentiment Analysis of Teachers’ Evaluation,” *Applied Computational Intelligence and Soft Computing*, Vol. 2016, Article 2385429 (2016).

- [8] Misuraca, M., Scepi, G., and Spano, M.: “Using Opinion Mining as an Educational Analytic: An Integrated Strategy for the Analysis of Students’ Feedback,” *Studies in Educational Evaluation*, Vol. 68, Article 100979 (2021).
- [9] Hujala, M., Knutas, A., Hynninen, T., and Arminen, H.: “Improving the Quality of Teaching by Utilising Written Student Feedback: A Streamlined Process,” *Computers & Education*, Vol. 157, Article 103965 (2020).
- [10] Santhanam, E., Lynch, B., and Jones, J.: “Making Sense of Student Feedback Using Text Analysis – Adapting and Expanding a Common Lexicon,” *Quality Assurance in Education*, Vol. 26, No. 1, pp. 60–69 (2018).
- [11] Sindhu, I., Daudpota, S. M., Badar, K., Bakhtyar, M., Baber, J., and Nurunnabi, M.: “Aspect-Based Opinion Mining on Student’s Feedback for Faculty Teaching Performance Evaluation,” *IEEE Access*, Vol. 7, pp. 108729–108741 (2019).