

要件定義書（PoC開発用）

プロジェクト名：オンライン授業コメント分類・分析システムPoC

開発期間：1ヶ月

目的：オンライン授業中に収集された受講者コメント（CSV形式）を自動分類・分析し、改善に役立つ洞察を抽出・可視化する仕組みを検証する。

1. 想定利用フロー

ステップ	処理内容
① CSVアップロード	授業コメントCSVを管理画面からアップロード
② コメント分類	コメントに対してカテゴリ・危険性・感情の自動ラベル付け（LLMまたは軽量モデルを使用）
③ コメント集約	感情別にコメントをクラスタリングして代表文を抽出（埋め込み＋クラスタリング）
④ 重要度評価	「緊急性×（質問＋インフラ＋具体性）」でスコア化し、重要コメントを抽出
⑤ データ可視化	PN比・カテゴリ別比率を円グラフで表示、重要コメントをランキング形式で表示

2. 機能要件（MVPレベル）

A. データ入力・管理

ID	機能名	説明
A1	CSVアップロード	コメントのCSVファイルをWeb上からアップロード
A2	コメント保存	コメントデータをDBに保存（原文＋各種ラベル用フィールド含む）

B. 自動ラベル付け

ID	機能名	説明
B1	カテゴリ分類	LLM or 軽量モデルで以下4分類を行う - 講義内容 / 授業資料 / 運営 / その他
B2	危険性判定	攻撃的・ハラスメント・不適切コメントの検出
B3	感情分類	ポジティブ / ネガティブ に分類 (0/1ラベル)

C. 集約・重要度評価

ID	機能名	説明
C1	感情別クラスタリング	ポジ/ネガ別に埋め込み計算 + クラスタリング (例: HDBSCAN) で代表文抽出
C2	タグラベル付け	各クラスタに以下のタグ (ワンホット) をLLMで付与 - 質問 / 具体的 / インフラ / 緊急性
C3	重要度スコア算出	重要度 = 緊急性 × (質問 + インフラ + 具体的) でスコア化

D. データ可視化 (UI設計含む)

ID	機能名	説明
D1	PN比グラフ	全体およびカテゴリ別のポジ/ネガ比率の円グラフ表示
D2	重要度ランキング表示	高重要度クラスタをランキング形式で表示、代表文とスコア付き
D3	コメント詳細表示	各クラスタ内コメントの閲覧機能 (代表文 + コメント例)

3. 非機能要件 (PoCレベル)

- **対象CSV形式**：1列目がコメント文、その他は無視。最大1000件程度想定。
- **LLM利用**：OpenAI API or Claude API (1つのリクエストでカテゴリ・感情・危険性を一括で返すプロンプト設計)
- **埋め込み+クラスタリング**：Sentence-BERT or Huggingfaceモデル (ローカルorAPI)
- **開発技術 (想定)**：
 - フロントエンド：React + Chart.js
 - バックエンド：FastAPI / Flask

- 。 データベース：SQLiteまたはPostgreSQL
- 。 分析処理：Python（pandas, scikit-learn, sentence-transformers）

4. スケジュール（1ヶ月目安）

週	タスク
1週目	要件定義精緻化、CSVアップロード画面作成、DB設計
2週目	LLMプロンプト開発、ラベル付け処理実装、感情分類・クラスタリング実装
3週目	タグ付け&重要度算出実装、グラフ表示、ランキングUI作成
4週目	テストデータで動作確認、UI調整、PoCデモ作成・改善反映

5. 今後の拡張見込み（PoC終了後）

- 。 単語・語彙頻度分析や時系列変化のダッシュボード化
- 。 重みを動的に変更できるUI
- 。 対応履歴管理との連携
- 。 マルチ授業・マルチCSV対応

6. リスクと対応

リスク	対応策
LLMコストが高い	コメント数制限＋一括処理で抑制、後に軽量モデルに切替可能性を検討
クラスタリングの品質	類似度閾値を手動調整できる管理機能を後に導入可能
スケジュール遅延	重要度ランキング機能を後回しにすることで回避可能

仕様書 | オンライン授業コメント分類・分析アプリ（PoC）

概要

本アプリケーションは、CSV形式で取得したオンライン授業中の受講者コメントを自動分類・分析し、視覚的に出力するPoCシステムである。主に以下の機能を備える：

- コメントのカテゴリ・感情・危険度の分類
- 感情ごとの代表コメント抽出とタグ分類
- 重要度スコアによる優先度ランキング
- グラフやリストによる可視化

1. CSVアップロード機能

項目	内容
機能名	CSVコメントデータのアップロード
入力	CSVファイル（UTF-8、1列目：コメントテキスト）
処理	コメントを1行ずつ読み取り、DBへ登録
出力	成功／失敗メッセージ
例外処理	空ファイル、不正なフォーマット（列数不一致）はリジェクト

2. 自動ラベル付け機能（LLM）

2.1 カテゴリ分類

項目	内容
分類対象	コメントテキスト
ラベル	講義内容 / 授業資料 / 運営 / その他
実装	OpenAI GPT-4 or Claude でプロンプト分類（JSON形式で返却）

2.2 危険コメント判定

項目	内容
ラベル	safe / dangerous
判定基準	攻撃的、ハラスメント、暴言などを含む内容
実装	同上。プロンプトで明示的に定義した上で分類

2.3 感情分類

項目	内容
ラベル	<code>positive</code> = 1 / <code>negative</code> = 0
判定基準	ポジティブな表現かネガティブな表現か
実装	同一プロンプト内で3分類をまとめて取得（API効率化）

3. 感情別クラスタリング機能

項目	内容
対象	分類済みのコメント（positive / negative）
技術	<code>sentence-transformers</code> による埋め込み+HDBSCANクラスタリング
出力	クラスタごとの代表文（中心文 or 最大密度文）と構成コメント
閾値	最小クラスタサイズを5〜10件に設定（設定可能に）

4. コメントタグ付け・重要度算出

4.1 タグ付け（ワンホット）

タグ名	意味	値
<code>質問</code>	質問・疑問点の提示	0 / 1
<code>具体的</code>	具体的な改善提案や事例	0 / 1
<code>インフラ</code>	通信・マイク・カメラなどの技術的問題	0 / 1
<code>緊急性</code>	今すぐ対処すべき内容か	0〜3（数値）

4.2 重要度スコア算出式

重要度 = 緊急性 × (質問 + 具体的 + インフラ)

例	計算
緊急性=3、質問=1、インフラ=1、具体的=0	$3 \times (1+1+0) = 6$

5. 可視化UI仕様（簡易設計）

5.1 PN比円グラフ

| 内容 | 全体コメントに対する positive / negative 比率を円グラフで表示 |

| 技術 | Chart.js または Recharts |

| 備考 | カテゴリ別PN比も同様の構成 |

5.2 重要度ランキング

| 内容 | 重要度スコアの高い順にクラスタ代表コメントを表示 |

| 表示項目 | スコア / コメント内容 / タグ一覧 / コメント件数 |

| オプション | 「コメントを見る」ボタンで代表クラスタ内のコメント一覧を展開可能にするUI |

6. データベース構成（簡易）

テーブル： **comments**

カラム名	型	説明
id	INTEGER	主キー
text	TEXT	コメント原文
category	TEXT	講義内容 / 資料 / 運営 / その他
danger	BOOLEAN	危険性判定結果
sentiment	INTEGER	感情分類 (1=positive, 0=negative)
embedding	VECTOR	文ベクトル
cluster_id	INTEGER	クラスタ番号
tags	JSON	タグ (質問, 具体的, インフラ, 緊急性)
importance_score	FLOAT	重要度スコア

7. 使用技術まとめ（PoC）

要素	技術
フロント	React / Next.js + Chart.js (グラフ)
バックエンド	FastAPI または Flask
LLM API	OpenAI GPT-4 or Claude
埋め込み・クラスタリング	sentence-transformers + HDBSCAN
DB	SQLite or PostgreSQL

その他	pandas, scikit-learn, UMAP(optional)
-----	--------------------------------------

8. テスト仕様（PoC用）

テスト内容	項目
入力ファイル確認	正常CSV・異常CSVでの動作確認
分類精度確認	各分類結果の妥当性を目視で検証（10件ずつ）
クラスタ代表確認	クラスタ内コメントと代表文の対応チェック
グラフUI確認	表示崩れ、数値整合性など