データサイエンスプロジェクト提案

[高度演習\_グループ5.pptx](https://docs.google.com/presentation/d/1_KaNus9ZP62f4Bney95P5BtMIuy3JmoJ/edit#slide=id.p3)

1. 班番号、メンバー名＆学籍番号

グループ5

NE21-1046C 渡邉陽太

NE21-1079G 兼子拳太朗

NE21-1084B 笹川高聖

NE21-1204A 志田優哉

NE21-1248J 吉田武琉

1. プロジェクトタイトル

DVを少なくしようプロジェクト

1. プロジェクト概要

本プロジェクトでは、もともとはDVが発生した家庭とその近隣地域に焦点を当て、DVと関連のある要素を分析によって抽出することを目的としていたが、家庭内暴力のようなプライベートな環境での情報収集に課題があったため、

暴力を未然に防いだり、早期発見するための施策を考察する。また、スマートスピーカーやスマートウォッチ等のデバイスが記録している家庭内音声を用いて機械学習を行い、音声からDVの有無を判別可能なシステムの考案も行う。

このシステムを有効活用できる範囲としては、保育園や学校、老人ホームや病院などが挙げられ、暴力が問題になってしまっている環境での対策として使用できればと考えている。

1. プロジェクト背景

新型コロナウイルス蔓延による外出自粛や在宅勤務などの急激な環境の変化によって、DVが急増。これらの背景から、私たちはデータサイエンスを用いて、DVの早期発見と防止、再発防止の施策及び、機械学習を用いたDV検出rシステムの考案の、マクロ・ミクロ両方の視点からこのプロジェクトを遂行していく。

・なぜデータサイエンスが重要なのか？

データサイエンスを用いることで、より高度な予測が可能となり、事前に兆候に気づける確率を高められると考えられる。また、人間が実際に常駐して調査しなくとも、スマートデバイスの活用により調査が可能となるため。

1. プロジェクト内容

4-1.プロジェクトの対象

・DVによって、被害を受けている女性・男性。

結果的に、療育施設や警察が活動する際に、その防止や対策業務の省力化、効率化を図ることができる。

4‐2.プロジェクトの目的

・　DVを減少させる。DV事案の予兆発見、未然予防に繋げる。

・　療育施設、警察のコストが減り、効率化を図る。

4-3．プロジェクトのアプローチ

* データ種類と収集方法  
  データの種類（データ形式、データ期間も合わせて）：

| **データの種類** | **データ形式** | **データ期間** | **データの収集方法** |
| --- | --- | --- | --- |
| **世帯年収** | **量的データ** | **過去6年**  **・コロナ前**  **・コロナ後** | **警察や児童相談所から**  **データをいただく** |
| **年齢** | **量的データ** | **警察や児童相談所から**  **データをいただく** |
| **性別** | **質的データ** | **警察や児童相談所から**  **データをいただく** |
| **地域住民との関係値** | **質的データ** | **警察や児童相談所から**  **データをいただく** |
| **DVの被害経験** | **質的データ** |  | **警察や児童相談所から**  **データをいただく** |

| データの種類 | データ形式 | データ期間 | データの収集方法 |
| --- | --- | --- | --- |
| 家庭内音声データ | 波形データ | 過去6年（コロナ前とコロナ禍） | スマートスピーカー・スマートウォッチ |
| 怒号リスト | 質的データ | 過去10年 | これまでの児童相談所から得られた暴言や怒号などのセリフを収集 |
| 怒号モデル | 波形データ | ー | これまでの児童相談所から得られた暴言や怒号などのセリフを元に、大声で発生した場合の波形を抽出 |
| 泣き声モデル | 波形データ | ー | これまでの児童相談所から得られた暴言や怒号などのセリフを元に、大声で発生した場合の波形を抽出 |

データの収集方法：

警察・児童相談所より、調書にある記載内容を教えていただく

ホームスピーカー（アマゾンecho,Google スピーカーなど)

ウェアラブルデバイス

周波数

家庭内の音（物音・叫び音・破壊音）

しかし、DV学習用音声データを実際に収集するのは難しそうなので、擬似的なサンプルで代用することを検討している。擬似的なサンプルとは、男性や女性のスタッフがさまざまな発言に対して演技を行い、擬似的なデータを収集する。

泣き声データについては保育園などに協力を仰ぎ、擬似的なデータを収集する。

＜データ分析手法＞  
→　収集したデータを用いて、どのような分析手法を採用して実行するかを説明すること。

・二項ロジスティック回帰分析

→目的変数をDVあり、なしの2つのカテゴリであるとし、説明変数としては、世帯年収、年齢および性別、地域住民との関係、DV被害経験などが挙げられる。そして、二項ロジスティック回帰分析を行うことで、どのような要因がDVを行う確率に影響を与えるのかを考察する。

・機械学習

モデルを構築・訓練するための波形データを収集し、データ作成及び特徴量を抽出し、適切な学習法を用いて訓練データを作成し、精度を評価するために検証データでテストを行う。そして、訓練・テストされたモデルを元に新規データ（リアルタイムによる家庭内音声データ）について推論を行う。

５．プロジェクト示唆

データ分析

・DVの予兆がある世帯を特定し、児童相談所や警察が目をかけておくときの参考にする。

・DVの予兆のある世帯の傾向を算出し、それらの世帯に対してDV防止施策を打つ

音声認識

・DV発生時の警察呼び出しによる早期対処

・警察への証拠提出及び捜査の補助的役割

６．プロジェクト懸念点

・説明変数のうち質的変数のカテゴリー分けが明確な定義のもとで行われていないと、正確な結果を導くことができない。

・どちらの分析手法にしても、データの偏りが生じないための適切なサンプリングと、より多くのデータが必要となる。

・家庭用スマートスピーカーの場合は基本的に常時録音がなされていないので、設置する場合、その費用の負担先であったり、常時録音などの特殊な設定が必要となる。

以上